MULTI-SCALE ANALYSIS TECHNIQUES IN PATTERN RECOGNITION SYSTEMS

by

Mudassir Masood

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

In Partial Fulfillment of the Requirements
for the Degree

MASTER OF SCIENCE

IN

Electrical Engineering

KING FAHD UNIVERSITY
OF PETROLEUM AND MINERALS

Dhahran, Saudi Arabia

April 2005

Dedicated to the most loving and precious to me, my mother Badar Jehan and my father Masood Ali, who always give the best of themselves to support me.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful

All praise and glory goes to Almighty Allah (Subhanahu Wa Ta'ala) who gave me the courage and patience to carry out this work. Peace and blessings of Allah be upon His last Prophet Muhammad (Sallulaho-Alaihe-Wassalam) and all his Sahaba (Razi-Allaho-Anhum) who devoted their lives towards the prosperity and spread of Islam.

First and foremost gratitude is due to the esteemed university, the **King Fahd University of Petroleum and Minerals**, and to its learned faculty members for imparting quality learning and knowledge with their valuable support and able guidance that has led my way through this point of undertaking my research work. My deep appreciation and heartfelt gratitude goes to my thesis advisor **Dr. Mohamed Deriche** for his constant endeavour, guidance and the numerous moments of attention he devoted throughout the course of this research work. He was always there to listen and to give advice. He taught me how to ask questions and express

my ideas. He showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. Working with him in a friendly and motivating environment was really a joyful and learning experience.

I extend my deepest gratitude to my thesis committee members Dr. Azzedine Zerguine and Dr. Umar Al-Suwailem who gave insightful comments and reviewed my work on a very short notice. I would also acknowledge Dr. Asrar for his cooperation in providing the TRL lab facility.

I would like to acknowledge Dr. Muhammad Hafiz Afzal for providing unconditional support, love and guidance to help us follow and understand the right path, our religion Islam. Its also due to him that I consider the period of my stay here in Saudi Arabia as the most important in my life. Acknowledgement is due to my senior fellows Muhammad Moinuddin, Saad Azhar and Sajid Anwar Khan for interesting and insightful discussions relating to my thesis and course work. Sincere friendship is the spice of life. I owe thanks to my house mates, colleagues and my friends for their help, motivation and pivotal support. A few of them are Juned Laiq, Shaikh Faisal Zaheer, Imran Azam, Khawar Khan, Saad bin Mansoor, Imran Naseem, Yasir Haroon, Aiman Rasheed, the two Adnans and many others; all of whom I will not be able to name here. They made my work and stay at KFUPM very pleasant and joyful. My heartfelt thanks to my days old friends Mehdi Muntazir, Muhammad Hasnain, Arslan Farrukh and Muhammad Fahd.

Last, but not least, I thank my family: my respected father, Masood Ali, and my

loving mother, Badar Jehan, for educating me, for unconditional love, support and encouragement to pursue my interests, even when the interests went beyond boundaries of language, field and geography. My sister Farheen Ahmer, for listening to me, and for believing in me. My brother Muhammad Ali, for answering all my e-mails and for all his precious time that he devoted so that I can stay in touch with my family through internet. He always reminded me that my research should always be useful and serve good purposes for all humankind.

May Allah help us in following Islam according to Quran and Sunnah! *(*Aameen)

Contents

List of Figures

List of Tables

THESIS ABSTRACT

Name: Mudassir Masood

Title: Multi-scale Analysis Techniques in Pattern Recognition Systems

Degree: MASTER OF SCIENCE

Major Field: Electrical Engineering

Date of Degree: April 2005

Good face detection is an important part of a reliable face recognition system. We propose methods of face detection and normalization. Normalization requires the exact location of eyes to straighten any rotated face image. Eyes detection methods are also developed for this purpose. To improve the accuracy of the classical PCA technique for face recognition, we developed a system based on the method of least-squares which takes into account the class information. We further improve accuracy by using multiple scales of the same face image. Such scales are obtained using the discrete wavelet transform. We show that there exists a particular set of subband images that are essential for robust face recognition. We propose here a model to find these subbands for input images. Based on this, we develop a multiscale face recognition system which utilizes the model to find important subbands. These subbands are then used for improving recognition accuracy.

Keywords: *PCA, MPCA, Discrete Wavelet Transform, Multiscale, Face Recognition, Face Detection, Eyes Detection.*

King Fahd University of Petroleum and Minerals, Dhahran.
April 2005

Chapter 1

Introduction

1.1 Motivation

With the advent of technology, the use of electronic devices has increased substantially during recent years. These devices are used in several applications to ease our life. The applications range from daily use of home appliances to some high performance scientific devices in modern labs. In many situations, the use of some of these devices must be restricted, to a limited group of people, for a number of reasons. For example, you will not allow anyone to access your computer and change the data without your permission. Similarly, it would be a big problem if anyone can access your bank account from an ATM. Some kind of key must be there so that only you, and not any stranger, can access the door or the account.

The most commonly used, and the earliest, technique is that of passwords. Pass-

words are very effective in the authentication of users. The user is required to type in his password on some kind of a keypad. If the password is correct the user is allowed access to the facility, otherwise the system rejects the user. There are several problems with the use of passwords. The passwords can be forgotten and even worse can be 'stolen'. To make it robust, cards were introduced in conjunction with passwords. The system can only be used for access by the person in possession of a card. Again, this poses several problems as the card can be lost, stolen or damaged or even duplicated. To avoid these troubles, systems based on biometrics started to gain more popularity and acceptance.

Biometrics is the way of getting a person's physical or behavioural characteristics or traits which can be used for authentication. Obviously, these features must be unique in order for the system to work. Biometrics help us avoid the problems discussed above. In fact, in biometrics-based systems, you are an identity by yourself. You are not dependent upon passwords and cards that can be lost and stolen. Several biometric systems exist today. They use different methods for authentication. Some typical biometrics used today include:

- Fingerprints
- Voice
- Face
- Hand Geometry

- Eyes Geometry
- Iris
- Ear Geometry
- Pattern of vein structure
  - in Retina
  - on the back of the hand
- Writing style
- Odor
- Gait
- DNA, etc.

The systems based on the above mentioned features have been shown to be very effective and efficient. Obviously, each has its own weakness which prevents it from being chosen as the best solution. For example, fingerprints are the most developed of these and achieve very good recognition accuracy but require human involvement (contact). This makes fingerprints non-user-friendly and unsuitable for many applications that prohibit human interaction with the system. Similarly, iris patterns have proved to be excellent features for human recognition, but the process of getting a good iris image is complex and expensive. If we look through the

list given above we find that most of these biometrics require some kind of subject involvement. Only biometrics based on facial features are capable of avoiding subject involvement. The face image can be easily captured by a camera remotely and non-invasively. The resulting image can be processed further to detect the exact face region and extract facial features. The importance of face recognition is further highlighted when we know that it can be used in fighting crimes. Surveillance cameras capturing images can be used to "keep an eye" on abnormal activities going around sensitive areas. The face images captured by these cameras at some crime scene can then be used to search through a database.

The face recognition problem can be defined as the process of identifying an individual from his/her face image. This face image can be captured by a camera or can be extracted from a video. Face recognition is a difficult process. This is mainly due to the similarities in face shapes of different people and the variations in face images of the same person. The image of a face is prone to change due to a number of factors like noise, illumination, viewpoint, age, expressions, occlusion etc. All these make face recognition a difficult task. In the past 30 years, the field of face recognition has witnessed major developments. The main reason for such expansion is the need for such systems for different commercial security applications. In spite of this advancement, the recognition systems have faced certain limitations in their effectiveness. Most systems will work only in a certain defined set of conditions. Two problems that are still greatly unsolved: the viewpoint or pose variations, and

illumination effects on face images. So it can be safely said that most face recognition systems lack robustness to different challenging conditions, and are still far away from human capabilities and from reaching maturity and widespread acceptance.

1.2 Objectives of the Research

The objective of this research is to design a face recognition system that is:

insensitive to pose variations, rotation, translation, and efficient in terms of computational complexity.

1.3 Contributions of the Thesis

In this thesis, a multiscale principal component analysis face recognition system is developed. The Wavelet transform is used for this purpose. It is shown that a particular set of subband images exists which is sufficient for robust face recognition. In addition, the problems of pose, including translation, rotation, and scale invariance are also tackled.

The main contributions of the thesis are:

- the development of a **pose invariant** technique for face recognition using the eyes images and Principal Component Analysis. (Sec. 5.3)

- the development of a class-dependent PCA for face recognition using the

method of Least-Squares. (Sec. 5.5)

- the development of a **face and eyes detection** method using Active Contour Models and Wavelet decomposition. (Sec. 4.3 & 4.4)
- the development of a technique for choosing the **best wavelet decomposition** for face images. (Sec. 6.6)
- the development of a new **multiscale PCA** technique for face recognition. (Sec. 6.7)
- the development of a new **Eyes database** to be used for face recognition using only the eyes, which will be made available to the research community everywhere. (Sec. 4.6)

1.4 Thesis Outline

The rest of the thesis is organized as follows. In chapter 2, a thorough literature review of the face and eyes detection and face recognition problems is given. All related technical material is explained in chapter 3. In chapter 4, the proposed methods for face and eyes detection are discussed and experimental results are presented. Face Recognition methods are discussed in chapter 5 followed by the proposed multiscale method for face recognition in chapter 6. In the end, chapter 7, a conclusion and future work directions are outlined.

Chapter 2

Background

Trying to imitate the best face recognition system of all, obviously the human brain, a number of studies were conducted in psychophysics. These studies aimed at understanding the way human brain works in image recognition. These studies in psychophysics suggest that humans use both holistic and local feature information for face recognition [3, 4]. If the local features are dominant enough to recognize a person, the holistic information may not be used. In machine recognition systems both approaches have been implemented. All face recognition methods obviously, before any recognition, first require that the facial structure is properly identified in the image.

2.1 Face Detection Research

Face detection from a single image continue to be a challenging task because of the variability in the pose, illumination, texture, and scale. Expressions, occlusions and size of the image also affect the detection process. Much work has been carried on face detection. A typical face detection algorithm must identify and locate faces in an image regardless of their position, scale, size, rotation, and illumination.

Several techniques have been used for this purpose. The earliest of these are the knowledge-based methods. In these, the human knowledge of face is translated into rules. The computer then uses these rules to find a face in an image. The most famous and the earliest work is the one by Yang and Huang [5]. They used a hierarchical rule-based method for face detection. In this method, a *multiscale* hierarchy of an image is calculated and a set of rules is applied on these images. Based on this work, Kotropoulos and Pitas [6] proposed their method that uses a projection method for facial feature location. The candidates gathered through projection, are then validated by applying some rules. In knowledge-based methods, human knowledge can not be translated into well-defined rules. Moreover most of these techniques can only work on frontal faces.

Techniques related to feature-based methods try to find those features that are invariant to common face image conditions like pose, illumination and size. Sirohey [7] proposed a localization method to segment a face from a cluttered background

for face identification. The method uses an edge map and heuristics so that only face edges are preserved. Instead of using edges, Chetrivikov [8] used blobs and streaks. The eyes, nose, and cheekbones were represented by blobs while the edges of eyes and lips were represented by lines. The image was then scanned to find occurrences of a combination of these particulars to a face image. Graf et al [9] developed a method that finds facial features from two binary images with the help of connected components. These binary images were constructed from processed gray-level images.

Based on random graph matching and local feature detectors, Leung et al [10] developed a probabilistic method for face detection. Their motivation is to formulate the face localization problem as a search problem in which the goal is to find the arrangement of certain facial features that is most likely to be a face pattern. Han et al [11] developed a face detection technique that uses eyes and eyebrows. They defined eye-analogue-segments as edges on the contours of eyes. The pixels at which the intensity value changes significantly were called eye-analogue-pixels and were used to generate segments that are helpful in the search for face regions.

Takacs et al [12] developed iconic filter banks, which are based on a biologically motivated image representation, for detecting facial landmarks. The filter banks are implemented via selforganizing feature maps (SOFM). The proposed method was a generic technique for object detection. However, before detection, an object (such as eye and mouth) model has to be trained. For an object with different orientations,

different models may be required. Human faces can be extracted from an image by using texture information as well. The work carried by Augusteijn and Skufca [13] for texture classification of human faces can be used to detect faces in complex images. The texture are computed using second order statistical features (SGLD)[14]. Dai et al [15] also applied the SGLD model to face detection. Color information was also incorporated with the face-texture model. Using the face texture model, Dai et al designed a scanning scheme for face detection in color scenes in which the orange-like parts including the face areas are enhanced. One advantage of this approach was that it could detect faces which were not upright or had special features such as beard and glasses.

Human skin colour is an important feature that was also used for face detection. A large number of methods are available that use skin colour information for face detection. Several studies have shown that the main difference in skin colours of different people lies in the intensity. The chrominance remains nearly the same [9, 16, 17]. Saxe et al[18] proposed an iterative skin identification method based on histogram intersection in the HSV colour space. Two histograms, an image histogram and the other control histogram (of a particular patch of skin) are calculated. Histogram intersection is used to compare the two histograms to decide upon a patch as a skin region or not. McKenna et al [19] presented an adaptive colour mixture model to track faces under varying illumination conditions. Instead of relying on a skin color model based on color constancy, they used a stochastic model to esti-

mate the object's color distribution online and adapt it to accommodate changes in the viewing and lighting conditions. Their results showed that they can track faces within a range of illumination conditions. However, this method cannot be applied to detect faces from a single image. Sobottka and Pitas [20] proposed a method for face localization and facial feature extraction using shape and color. Skin-like regions are extracted by colour segmentation in the HSV space. Connected components are then determined by region growing at a coarse resolution. A best fit ellipse is computed using geometric moments for each region. Connected components that are well approximated by an ellipse are then selected as face candidates.

In contrast to non-parametric methods, Gaussian density functions [21, 22] and mixture of Gaussians [23, 24, 25] have also been used to model skin regions. The Expectation Maximization (EM) algorithm is usually used to estimate the parameters of Gaussian mixture models [23, 24]. Hawkins' statistical test on the normality of the estimated Gaussian mixture models is performed and McLachlan's bootstrap method was used to test the number of components in a mixture. Their experimental results show that the estimated Gaussian mixture model fits skin images from a large database.

In template matching, a frontal face image template is predefined in the system. This template is then matched with an input image. The match is calculated as the correlation between the template and the input image. Some template matching techniques do not use such predefined template. Such techniques cannot handle

effectively the problems of scale, pose, and shape. Sakai et al [26] attempted to detect frontal faces by using subtemplates of eyes, nose, lips etc. They calculated the correlation between a particular subtemplate and the image. This signaled the region of interest. Once the region of interest is known other subtemplates are used to verify the presence of a face.

Craw et al [27] used Sobel operators to extract the face boundary. This extracted boundary is matched to a predefined template. After the head contour has been located, other detailed features of the eyes, and nose are located. A more advanced hierarchical template matching method for face detection was also proposed by Miao et al. [28]. First the input image is rotated from -20 degrees to 20 degrees in steps of $5°$, in order to handle rotated faces. A multiresolution image hierarchy is then formed and edges are extracted using the Laplacian operator. The face template consists of the edges produced by six facial components: two eyebrows, two eyes, one nose, and one mouth.

Finally, heuristics were also applied to determine the existence of a face. Their experimental results show better results in images containing a single face (frontal or rotated) than in images with multiple faces. The work of Yuille et al [29] in modeling facial features using deformable models is the first to tackle the modeling problem. In this approach, facial features are described by parameterized templates. An energy function is defined to link edges, peaks, and valleys in the input image to corresponding parameters in the template. The best fit of the elastic model is

found by minimizing an energy function of the parameters. One drawback of this approach is that the deformable template must be initialized in the proximity of the object of interest. A face detection method based on the use of snakes, also known as active contour models, was proposed in [30]. The input image is first blurred by a filter, and the edges enhanced by using morphological operations. A snake is then initiated for this processed image which then deforms to get the face boundary. The basic theory of snakes is similar to the deformable templates by Yuille. Lam and Yan [31] and Yuen et al. [2] employed a snake model for detecting face boundary. A snake gives a good result in boundary detection if the initial position of the snake contour is close to the target. Matsuno et al. [32] proposed a potential net to detect human face. The precise location of the face is determined by horizontal and vertical projections. Jeng et al. [33] adopted morphological operations and boost-filtering to remove the effect of complex background. Several other face detection techniques have also been developed with limited success. For details see [34] and [35].

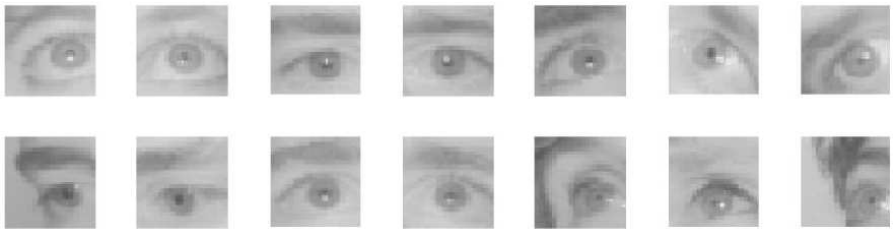
2.2 Eyes Detection Research

The detected human face in a scene can be further processed to locate the exact positions of the eyes. The position of eyes is important in the normalization process. Normalization is a necessary step for a number of face recognition algorithms. In normalization, the location of the eyes is used to rotate and scale face images. Several

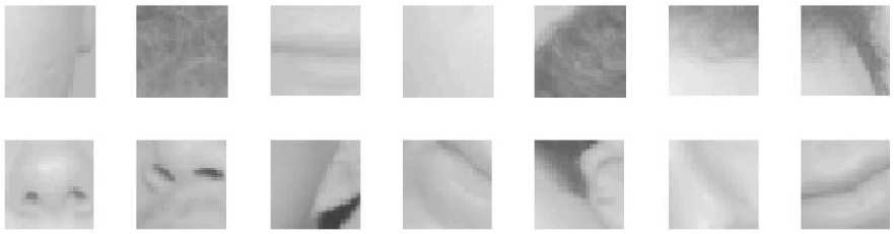
eyes detection algorithms have been developed in the literature. Most of these assume that the location of face is known. Once the face boundary is known, the rough eye windows can be located. Some algorithms use this information to increase their performance accuracy. Various methods have been proposed to describe the geometrical information of eye features, such as chain codes, fitting line segments, autoregressive models, Fourier descriptors, active contour models, and deformable templates to mention a few.

Yuille et al. [29] first proposed the use of a deformable template in locating human eyes. In this method, an eye model is designed and the eye position is obtained through a recursive process. However, this method is applicable only if the initial position of the eye model is placed near the actual eye position. Moreover, deformable template suffers from two limitations: first, it is computationally expensive and, second, the weighting factors for energy terms have to be determined manually. Improper selection of the weighting factors results in inaccurate localization. In view of these limitations, Lam et al. [36] introduced the concept of eye corners to guide the recursive process and partially solved the problems. In [36], Xie et al. [37], the corner detection algorithm was also adopted. The algorithm was based on the edge image, however, a good edge image is hard to obtain when the contrast of eye image is relatively low.

Some other methods have also been proposed using deformable templates to extract eye features, for example [38, 39, 40]. Similar to the Lam et al. approach,

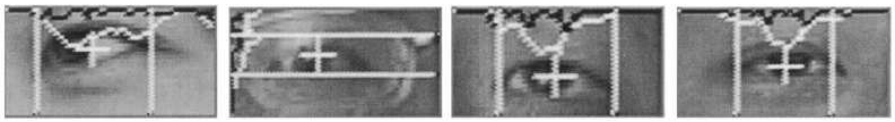


(a) Eye images

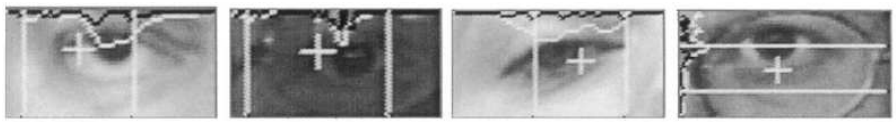


(b) Non-eye images

Figure 2.1: Example of eyes detection by Feng and Yuen [1]



(a) Correct detection of eye images



(b) Wrong detection of eye images

Figure 2.2: Example of eyes detection using different variance projection functions [2]

Feng and Yuen [2] developed a variance projection function (VPF) for locating the landmarks (corner points) of an eye. It was observed that some eye landmarks have relatively high contrast, such as the boundary points between eye white and eye ball. The located landmarks are then employed to guide the eye detection process. Saber et al. [41] and Jeng et al. [33] proposed to use the facial features' geometrical structure to estimate the location of eye whereas the precise locations have not been reported in their work. Feng and Yuen [1] proposed the use of an eye variance filter. The exact location of eyes is found by convolving the face image with this filter. The filter is formulated using some training images of the eyes. The technique does not work well if the image scale is changed.

2.3 Face Recognition Research

Face recognition techniques can be broadly divided into two classes: feature based, and holistic algorithms. The earliest systems for face recognition were feature-based. They used local features like distance between different parts of the face, for example distance between the two eyes, width of the head and the distance between eyes and mouth [42, 43]. Similarly, the angles between these features were also calculated to form feature vectors comprising these distances and angles. The first holistic face representation technique that used the complete face image as input is the *eigenpictures* approach developed by Sirovich and Kirby [44]. They used Principal

Component Analysis (PCA) also known as Karhunen-Loève (KL) transform for this purpose. A great amount of work has been done in both categories of face recognition techniques as witnessed during the last three decades. The above approaches are discussed in more details in the following sections.

2.3.1 Holistic Face Recognition Algorithms

Using the method of eigenpictures proposed by [44], Turk and Pentland [45] developed a face recognition technique called *eigenfaces*. PCA uses second order statistics of images and optimally minimizes the reconstruction error under the L^2 norm assuming Gaussian distribution. PCA is not optimized for class separability. To cope with pose variations and to utilize local feature information, view-based and modular eigenspace methods have been proposed respectively as extensions to the eigenfaces approach. In the view-based eigenspace method, for each different view of the individuals, an eigenspace is constructed. When a test image arrives, first its viewspace is estimated, then classification is done in that viewspace. This method performs better than the simple eigenface technique in which only one space is constructed and the different views of individuals are disregarded.

In modular eigenspace method, an eigenspace is constructed for each of the local facial features like eyes, nose, and lips. Combining local feature information with eigenfaces improves recognition rate. Some of the work by Pentland [46] and Sung [47] showed improved performance compared to simple PCA. Dual eigenface

technique is another extension of the ordinary eigenface approach. In this method, intra personal and extra personal differences are used to exploit the knowledge of critical variations. A probabilistic similarity measure based on Bayes rule is used for classification. When a test image arrives, it is subtracted from the face images in the database, and the obtained differences are both projected onto the intra personal and extra personal difference subspaces. The likelihoods in these subspaces are estimated to derive the maximum a posteriori (MAP) similarity measure. This method is found to result in superior performance compared to the eigenface method [44].

Principal Component Analysis is best suited in extracting linear relations between variables in a given dataset. It can be shown that if the data statistics are Gaussian, the PCA is the best method for dimension reduction [48, 49]. This is because PCA is purely a second-order method. However, if this is not the case, and the data statistics are, for example, mixture of Gaussians, then PCA does not perform well. A Nonlinear Principal Component Analysis (NLPCA) technique is used for such cases. The NLPCA has been functionally defined in [50] by a class of autoassociative neural networks. They proposed a 3-hidden layer neural network.

The eigenfaces (PCA) technique has a major drawback as it does not consider class information. Linear Discriminant Analysis (LDA) also known as Fisher Discriminant Analysis (FDA) [51] takes this into account. LDA performs better when the number of samples per class is large. On the other hand if the number of samples

per class is very small, PCA is a better choice. More detailed comparison of PCA and LDA can be found in Beveridge et al [52] and Martinez et al [53]. Another technique that uses higher order statistics and was shown to provide better recognition performance [54] but at higher computational load is the Independent Component Analysis (ICA).

PCA or the KL transform is the optimal linear transformation that is used for dimension reduction. Only the subspace with largest variances is kept. This means that it reduces the dimensionality of the dataset while maintaining those characteristics of the dataset that contribute most of the variance. Unlike other transforms such as the discrete cosine transform, the basis vectors are not fixed in PCA. In image recognition PCA is shown to preserve most of the power in signals. However it is very poor in discrimination, meaning that the class separability is not good.

The other problem with PCA is its high computational load. To overcome the problem of high computational cost, one can use only some portions of face instead of using the complete face image as input to the eigenpicture technique. These can contain images of portions like eyes, nose, lips etc. Experimental results show that comparable performance can be achieved by using only eyes from a face image [55]. The other alternative is to use small size images for the faces. However, for some methods that depend on feature localization on the face, small images will not work. One such example is the graph matching method by Okada et al [56]. However for

holistic face recognition, the image size can be very small. For example 12x11 for the subspace LDA system [57], 14x10 for the Probabilistic Decision-Based Neural Network (PDBNN) system [58], and 18x24 for human perception [59]. Zhao et al [57] postulated that there exists a particular subspace for face representation. This subspace is of fixed size. If any face representation exceeds this size there is not much affect on the overall recognition. On the other hand, a face representation less than the size of this subspace will degrade the system performance. Experiments were conducted with images of different sizes to obtain different face subspaces. As the SNR increases with decrease in image size, a better performance was observed when smaller face images were used [60].

As discussed above, higher SNR might be one of the reasons behind better performance, the experimental results verify that accurate feature location is also critical for good recognition performance. We note that holistic methods require images to be normalized according to some criteria based on the features location. The method commonly used is to normalize face image with respect to the location of eyes [60, 61]. In smaller images, the location error is slightly reduced hence resulting in some improvement in performance [58]. It should be noted here that very small size images make the recognition task very difficult. The normalization of face images is performed with respect to translation, in-plane rotation, and scale. This type of normalization was mainly used in [62, 63, 45, 64], whereas in [65] the normalization also included masking and affine warping to align the shape. In the work

by Craw and Cameron [66], manually selected points were used to warp the input images to the mean shape, yielding shape free images. Because of this difference, PCA was shown to be a good classifier in [65] for the shape free representations, but may or may not be as good for the simply normalized representations.

2.3.2 Feature-based Face Recognition

The use of the local features as discussed above is not new. A number of methods in this category have been developed. These include methods based on the geometry of local features [43, 42] as well as 1D and pseudo-2D HMM methods [67, 68]. One of the most successful techniques is the Elastic Bunch Graph Matching (EBGM) system [56, 69], which is based on the Dynamic Link Architecture (DLA) [70, 71]. Wavelets also played a major role in face representation in these graph matching methods. A typical local feature representation consists of wavelet coefficients for different scales and rotations based on fixed wavelet bases. These locally estimated wavelet coefficients are robust to illumination change, translation, distortion, rotation, and scaling.

Compared to holistic approaches, feature-based methods are less sensitive to variations in illumination and viewpoint and to inaccuracies in face localization. However the feature extraction techniques needed for this type of approach are still not reliable or accurate enough. For example, most eye localization techniques assume some geometric and textural models and do not work if the eye is closed.

2.3.3 Multiscale Techniques in Face Recognition

During the last few years we have witnessed a number of wavelet-based techniques proposed for face recognition. The systems that have been discussed so far for face recognition consider the input images at a single scale. This kind of single scale representation is appropriate if the input image contains contributions at just one scale. This is not true for face images as they are multiscale in nature. This is also not true for most data we encounter [72] because of the events occurring at different locations and with different localization in time and frequency. So, to analyze properly such data, we must consider multiscale modelling. One such example is [72] where multiscale PCA for modelling of data was used in multivariate statistical process monitoring.

In the case of images, especially faces, multiscale PCA was attempted by Feng et al. [73]. In [73], a primitive approach for face recognition was discussed. The authors selected a *subband* image on the basis of experiments they performed on a particular database. They claim that they could get best recognition rate using only a particular subband image. Their claim lack generality and is based on experiments on a single database. Another wavelet transform based identification system was proposed in [74]. In this system, a three level wavelet decomposition is performed. The scaling components at each level as well as the original image are used for classification. The classifier used in this study is a kind of neural network with one-

class-in-one-network structure. This means that each subnet is trained separately and there is one subnet per class. Higher classification rates were obtained with the proposed system compared to the standard eigenface method on ORL(ATT) database.

Garcia et al. [75] have used a two level wavelet packet decomposition of images to calculate a set of features. This decomposition is performed on the detected faces from images. For this purpose, a face bounding box is found. For each subband, a set of statistical features was calculated. These features include the means and the variances of the subband images. The Bhattacharyya distance was used to classify these feature vectors. An extension to their work is presented in [76] where they proposed another technique to find the face bounding box. However, the technique is highly dependent upon the nature of image and does not work if the images vary from the conditions specified in the training.

Perlibakas [77] used a wavelet packet decomposed images with principal component analysis. Their aim was to speed up the pca-based recognition process. They showed that their technique only works if more than 650 images are present in a certain database. Their results are based on a specific database they used and the results can not be generalized. Their results show that they were unable to get improvement in the recognition rate. In [78] a three level wavelet decomposition was proposed and the resulting approximation subbands at each level are concatenated to produce a new data vector. PCA is then applied on this new data vector.

Radial basis functions are used as the classifier of the system. The proposed system achieves higher correct recognition rate than the standard eigenface method on the ATT database.

A Discriminant *waveletfaces* approach was proposed in [79]. In that study, a three level wavelet transform was applied on the original image and the resulting third level approximation subband, waveletface, is used as the input of the system. LDA is performed on this waveletface. For classification, the nearest feature plane (NFP) and nearest feature space (NFS) classifiers are discussed. The proposed method was tested on the IIS and ATT databases. From the experimental results, it was found that discriminant waveletfaces combined with the NFS classifier achieve the highest correct recognition rates. The technical literature available for the subject of face recognition using multiscale techniques is still very limited as not much work has been done in this field. Most of the work is based on trial and error with no proper formulation. In view of this limited work, the proposed work of this thesis is to build some formal structure to utilize the multiscale properties of face images. This is discussed in detail in Chapter 6.

2.3.4 Hybrid Techniques in Face Recognition

In Hybrid approaches, both holistic and local features are used. In [80], a modular eigenfaces approach was proposed where both global eigenfaces and local *eigenfeatures* are used. Note that the concept of eigenfaces can also be extended to

eigenfeatures, such as *eigeneyes, eigenmouths*, etc. Using a limited set of images (45 persons, two views per person, with different facial expressions such as neutral vs. smiling), recognition performance as a function of the number of eigenvectors was measured for eigenfaces only and for the combined representation. For lower-order spaces, the eigenfeatures performed better than the eigenfaces [80]. When the combined set was used, only slight improvement was observed.

The view that hybrid systems should be used for face recognition is supported by many researchers and the psychology community [81]. In [82], it was suggested that a hybrid of PCA and local feature analysis (LFA) can be used. LFA is an interesting biologically inspired feature analysis method [82]. Its biological motivation comes from the fact that, though a huge array of receptors (more than six million cones) exist in the human retina, only a small fraction of these is active, corresponding to natural objects that are statistically redundant [83]. From the activity of these sparsely distributed receptors, the brain has to discover where and what objects are in the field of view and recover their attributes. Consequently, one expects to represent the natural objects in a subspace of lower dimensionality by finding a suitable parameterization. For a limited class of objects such as faces which are correctly aligned and scaled, this suggests that even lower dimensionality can be expected [82]. One good example is the successful use of the truncated PCA expansion to approximate the frontal face images in a linear subspace [44, 84].

Another method proposed by [85] that uses hybrid technique was based on recent

advances in component-based recognition [86] and 3D morphable models [87, 88]. The basic idea of component-based methods [86] is to decompose a face into a set of facial components such as mouth and eyes that are interconnected by a flexible geometrical model. The motivation for using components is that changes in head pose mainly lead to changes in the positions of facial components which could be accounted for by the flexibility of the geometric model. However, a major drawback of the system is that it needs a large number of training images taken from different viewpoints and under different lighting conditions. To overcome this problem, the 3D morphable face model [88] was applied to generate arbitrary synthetic images under varying pose and illumination. Only three face images (frontal, semiprofile, profile) of a person are needed to compute the 3D face model. Once the 3D model is constructed, synthetic images of different sizes are generated for training both the detector and the classifier.

In this study we focus on finding a face recognition method on the basis of frequency content of input images. This method will be pose, rotation, translation and scale invariant.

2.4 Summary

In this chapter a detailed review of the literature was presented. We discussed the methods and techniques of the major developments in the fields of face and eyes

detection, and recognition. Special emphasis was given to the different techniques of face recognition including the multiscale techniques. In the next chapter, we will present some of the fundamental concepts necessary to discuss our proposed algorithms. This will include a detailed explanation of the mathematical techniques used in our research.

Chapter 3

Fundamental Concepts

3.1 Overview of Principal Component Analysis (PCA)

Principal Component Analysis also known as the Karhunen-Loève transform is a mathematical technique used for dimension reduction while preserving the components that contribute most to the variations in the original data. This process transforms a number of correlated variables into a smaller (or equal) number of uncorrelated variables. The first principal component contribute most to the variance and is a strong indicator of the relation between variables. The following principal components also indicate the dependency among the variables but are successively of lesser importance.

The process of PCA is explained in detail in the next section. In order to have a good understanding of PCA we would briefly mention that the process of PCA involves calculation of eigenvalues and eigenvectors of a covariance matrix. These eigenvectors are the principal components of the data. If we arrange the eigenvectors such that their corresponding eigenvalues are in descending order, the first eigenvector will be the most significant. Similarly, the eigenvector corresponding to the smallest eigenvalue will be least important. Since the technique is simply a mathematical transform, we can recreate the original data from these eigenvectors. In the process of recreating the data and if we do not consider some of the lesser important components, the new generated data will be of reduced dimension. We will lose some information, but such loss will not affect the total information content of the data. Fig. 3.1 shows the energy information of eigenvectors of an example data set. We can see that most of the energy is concentrated in the first few eigenvectors.

The field of pattern recognition usually involves high-dimensional data and it is very difficult to extract similarities and patterns in such a data. PCA has found its application in the field of pattern recognition because of two main reasons:

1. the ability to find patterns in high dimensional data
2. the ability to reduce the dimensionality of data with a minimal loss of information.

The goal of such approach in pattern recognition is to represent a picture with as

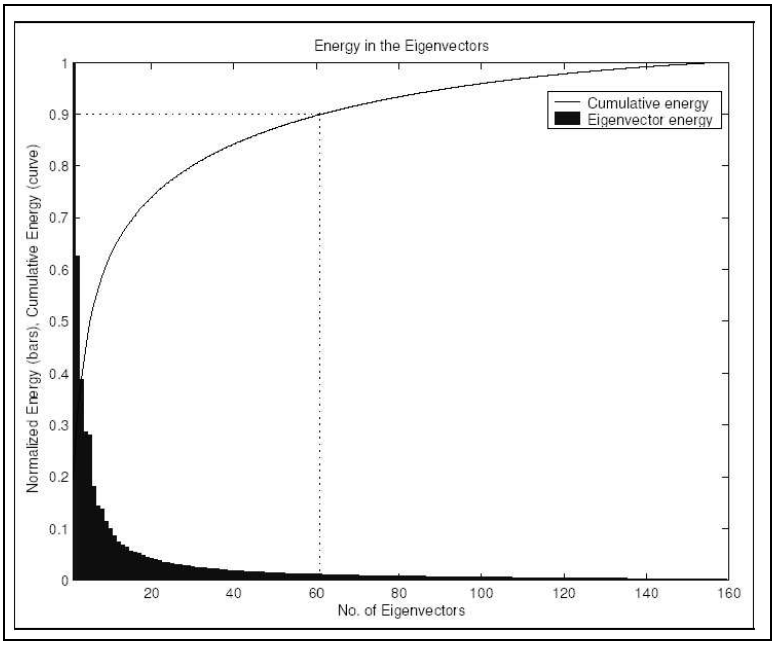


Figure 3.1: Eigenvectors energy details

few eigenvectors as possible such that the mean-square error introduced by ignoring the less significant eigenvectors is minimum. To achieve this, the covariance matrix of the ensemble of pictures is used to find the eigenvectors. We call these eigenvectors *eigenfaces* when working with face images.

3.1.1 Mathematical Description

A given $N \times N$ face image X_k can be represented as a vector $\mathbf{x}_k$ in a number of ways. The easiest and simplest method is to select the first row X_{1j} of the image and use it as the first N components of the vector $\mathbf{x}_k$. Similarly the second row X_{2j} will be concatenated to vector $\mathbf{x}_k$. In this manner we create vector $\mathbf{x}_k$ which is shown in equation 3.1.

Note that in this method we are relating the far ends of each two successive rows. These two ends do not have any correlation. This can affect the performance of PCA because PCA looks for the correlation between adjacent elements of the vector $\mathbf{x}_k$. In order to reduce this effect we tried another method called *circular sampling*. In circular sampling, we start from the center of the image and select pixels in a circular fashion. In this way we are able to formulate vector such that there is always correlation between adjacent pixels. However, to be able to compare our results to existing ones, we opted to use the raster scanning approach based on the row by row concatination.

Lets say now that we are concerned with a set of M images, X_k, $(k = 1, ..., M)$. Each image is of size $N \times N$. Each image X_k is expressed in the form of an N^2-dimensional column vector $\mathbf{x}_k$ as follows [89]:

$$\mathbf{x}_k = [x_{k1} \quad x_{k2} \quad \dots \quad x_{kj} \quad \dots \quad x_{kN^2}]' \tag{3.1}$$

The vectors $\mathbf{x}_k$ where $k = 1, ..., M$ can be seen as M observation vectors of a certain random vector $\mathbf{x}$. The covariance matrix of such random vector $\mathbf{x}$ is defined as:

$$\mathbf{C}_\mathbf{x} = E\{(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{x} - \mathbf{m}_\mathbf{x})'\} \tag{3.2}$$

where

$$\mathbf{m}_\mathbf{x} = E\{\mathbf{x}\} \tag{3.3}$$

is the mean vector, E is the expected value, and the prime($'$) indicates transposition. Equations (3.2) and (3.3) can be approximated from the M observations using the

following:

$$\mathbf{m_x} \cong \frac{1}{M}\sum_{k=1}^{M}\mathbf{x}_k \tag{3.4}$$

and

$$\mathbf{C_x} \cong \frac{1}{M}\sum_{k=1}^{M}(\mathbf{x}_k - \mathbf{m_x})(\mathbf{x}_k - \mathbf{m_x})' \tag{3.5}$$

or equivalently

$$\mathbf{C_x} \cong \frac{1}{M}\left[\sum_{k=1}^{M}\mathbf{x}_k\mathbf{x}_k'\right] - \mathbf{m_x}\mathbf{m_x}' \tag{3.6}$$

Now, in order to find such a transformation we start by decomposing the covariance matrix $\mathbf{C_x}$ using the Singular Value Decomposition. Starting with the mean vector and the covariance matrix, PCA aims at obtaining a linear transformation matrix $\mathbf{A}$. This matrix can be used to transform the random vector $\mathbf{x}$ of dimension N^2 into a random vector $\mathbf{y}$ of dimension $K << N^2$. Let $\mathbf{e}_i$ and λ_i be the eigenvector and corresponding eigenvalue of $\mathbf{C_x}$, $1 \leq i \leq M$. Then, it can be shown that a transformation matrix whose rows are the eigenvectors of $\mathbf{C_x}$ can be used to uncorrelate the elements of $\mathbf{x}$.

$$\mathbf{A} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1N^2} \\ e_{21} & e_{22} & \dots & e_{2N^2} \\ \vdots & & & \\ e_{N^21} & e_{N^22} & \dots & e_{N^2N^2} \end{bmatrix} \tag{3.7}$$

The new random vector $\mathbf{y}$ can hence be obtained using the following transform:

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m_x}) \tag{3.8}$$

Note that the covariance matrix of $\mathbf{y}$ is given by:

$$\mathbf{C_y} = E\{(\mathbf{y} - \mathbf{m_y})(\mathbf{y} - \mathbf{m_y})'\} \tag{3.9}$$

we can also show that $\mathbf{m_y} = \mathbf{0}$

$$\begin{aligned}
\mathbf{m_y} =& E\{\mathbf{y}\} \\
=& E\{\mathbf{A}(\mathbf{x} - \mathbf{m_x})\} \\
=& \mathbf{A}E\{\mathbf{x}\} - \mathbf{A}\mathbf{m_x} \\
=& \mathbf{0}
\end{aligned} \tag{3.10}$$

The substitution of equation (3.8) and (3.10) into equation (3.9) yields the following expressions for $\mathbf{C_y}$ give

$$\begin{aligned}
\mathbf{C_y} =& E\{(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m_x})(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m_x})'\} \\
=& E\{\mathbf{A}(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})'\mathbf{A}'\} \\
=& \mathbf{A}E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})'\}\mathbf{A}' \\
=& \mathbf{A}\mathbf{C_x}\mathbf{A}'
\end{aligned} \tag{3.11}$$

Since $\mathbf{C_x}$ is a real, symmetric matrix it is always possible to find a set of orthonormal eigenvectors [90] and the matrix can be decomposed into factors $\mathbf{C_x} = \mathbf{A^{-1}}\mathbf{\Lambda}\mathbf{A} = \mathbf{A}'\mathbf{\Lambda}\mathbf{A}$. $\mathbf{\Lambda}$ is the diagonal matrix composed of the eigenvalues of $\mathbf{C_x}$.

Hence equation (3.11) can be further simplified:

$$\begin{aligned}\mathbf{C_y} =& \mathbf{A}\mathbf{C_x}\mathbf{A}' \\ =& \mathbf{A}\mathbf{A}'\mathbf{\Lambda}\mathbf{A}\mathbf{A}' \\ =& \mathbf{I}\mathbf{\Lambda}\mathbf{I} \\ =& \mathbf{\Lambda}\end{aligned} \tag{3.12}$$

This shows that $\mathbf{C_y}$ is a diagonal matrix with elements equal to the eigenvalues of $\mathbf{C_x}$, that is

$$\mathbf{C_y} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & \ddots & \\ 0 & & \lambda_{N^2} \end{bmatrix} \tag{3.13}$$

This means that the elements of $\mathbf{y}$ are uncorrelated.

$\mathbf{x}$ can be reconstructed from $\mathbf{y}$ using the relation

$$\mathbf{x} = \mathbf{A}'\mathbf{y} + \mathbf{m_x} \tag{3.14}$$

If we form $\mathbf{A_K}$ from only K eigenvectors corresponding to the largest eigenvalues, the random vector $\mathbf{y}$ will be K-dimensional and $\mathbf{x}$ can be approximated as

$$\hat{\mathbf{x}} = \mathbf{A}_K'\mathbf{y} + \mathbf{m_x} \tag{3.15}$$

As such, by selecting the number of eigenvectors (K), we can perform the dimensionality reduction according to our requirements mentioned earlier.

3.1.2 Reduced Computational Complexity

It can be seen that the size of $\mathbf{C_x}$ is too large $N^2 \times N^2$(N being the number of pixels in an image), even for small images. This means that the computation of eigenvectors is complicated (if $N^2 > M$). Sirovich and Kirby [44] proposed a decomposition method to reduce the dimensionality of matrix $\mathbf{C_x}$ and hence the overall computational complexity. According to this method,

$$\mathbf{C_x} = E\{(\mathbf{x} - \mathbf{m_x})'(\mathbf{x} - \mathbf{m_x})\} \tag{3.16}$$

which results in $\mathbf{C_x}$ of size $M \times M$, where M is the number of observations of the random vector $\mathbf{x}$. According to the method proposed by Sirovich and Kirby the size of the new transformation matrix $\mathbf{A}$ becomes $M \times N^2$.

$$\mathbf{A} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1N^2} \\ e_{21} & e_{22} & \dots & e_{2N^2} \\ \vdots & & & \\ e_{M1} & e_{M2} & \dots & e_{MN^2} \end{bmatrix} \tag{3.17}$$

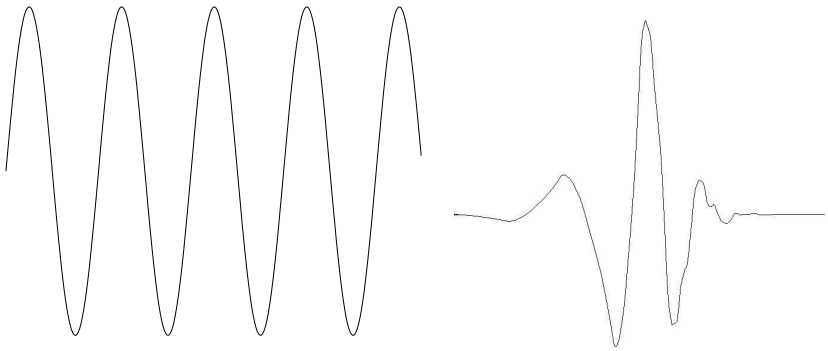
This result suggests that we can not use more than M eigenvectors for image representation. Usually for any practical purpose database we do not need more than M eigenvectors for good results. Often the task can be performed by much less eigenvectors than M.

3.2 Overview of Wavelet Theory

3.2.1 Introduction to the Wavelet Transform

The wavelet transform has been introduced in the literature as a transformation over basis functions that have attractive properties in both time and frequency domains. These basis functions are called wavelets. The wavelet transform is similar to the Fourier transform in the sense that basis functions are localized in frequency. In Fourier transform, we do not have temporal localization i.e the basis functions have infinite length. While in wavelet transform we have temporal or spatial localization, this means that wavelets are limited size functions. Figs. 3.2(a) and 3.2(b) show the basis functions used for the Fourier transform and the wavelet transform, respectively. In order to understand clearly the concept of temporal and spatial localization see Fig. 3.3.

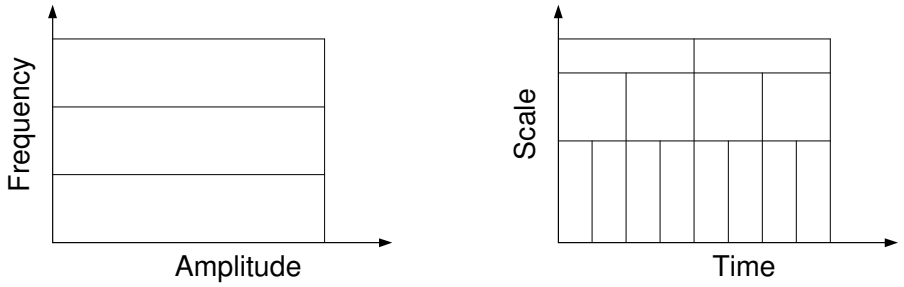
In the last years, wavelets have become very popular with the main reason being the development of a good framework for wavelet bases construction and efficient algorithms for the computation of such transform[91, 92].



(a) Basis for Fourier transform

(b) Sample basis for Wavelet transform (db4)

Figure 3.2: Comparison of the bases for Fourier and wavelet transforms



(a) Fourier transform

(b) Wavelet transform

Figure 3.3: Windowing for Fourier and wavelet transforms

The Continuous Wavelet Transform

The continuous wavelet transform of a square integrable function (finite energy signal) $s(t)$ is defined as [93]

$$C(a,b) = \int_R s(t) \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) dt, \qquad a \in R^+ - \{0\}, b \in R, s \in L^2(R) \tag{3.18}$$

where $C(a,b)$ are the continuous wavelet transform coefficients. The function $\psi(.)$ is called the mother wavelet while a and b are the parameters to control respectively the scale and position of the mother wavelet. The mother wavelet $\psi(.)$ has to satisfy the admissibilty criterion to ensure that it is a localized zero mean function. Mathematically, the admissibility criterion is given as follows.

$$\int_{R^-} \frac{|\psi(s)|^2}{|s|} ds = \int_{R^+} \frac{|\psi(s)|^2}{|s|} ds = K_\psi < +\infty \tag{3.19}$$

When the energy of the signal is finite, not all values of a decomposition are needed to exactly reconstruct the original signal, provided that the wavelet used satisfies admissibility criterion. In such cases, discrete analysis is sufficient and continuous analysis becomes redundant.

The Discrete Wavelet Transform

Equation (3.18) can be discretized by restraining a and b to a discrete lattice ($a = 2^j, b = k2^j, (j,k) \in Z^2$).

The discrete wavelet transform of a square integrable function $s(t)$ is defined as

$$C(j,k) = \int_R s(t)\psi_{j,k}(t)dt, \qquad (j,k) \in Z^2, s \in L^2(R) \tag{3.20}$$

where $C(j,k)$ are the discrete wavelet transform coefficient. $\psi_{j,k}(t)$ are the wavelet expansion functions or the wavelet basis functions. These are related to the original mother wavelet function denoted as $\psi(.)$ and are given as follows.

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{3.21}$$

where j and k are the dilation and translation parameters respectively.

Inverse Wavelet Transform

Under the admissibility conditions, the inverse wavelet transform exists and is given respectively for both continuous and discrete cases below:

$$s(t) = \frac{1}{K_\psi}\int_{R^+}\int_R C(a,b)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\frac{da}{a^2}\frac{db}{} \tag{3.22}$$

where K_ψ is a constant depending on ψ.

$$s(t) = \sum_{j\in Z}\sum_{k\in Z} C(j,k)\psi_{j,k}(t) \tag{3.23}$$

Typically some more constraints are imposed on ψ to ensure that the transform is non-redundant, complete and constitutes a multiresolution representation of the original signal. This leads to an efficient real-space implementation of the transform

using quadrature mirror filters. A combination of low and highpass filters is used. The low and highpass filters at the decomposition level are shown in Fig. 3.4.

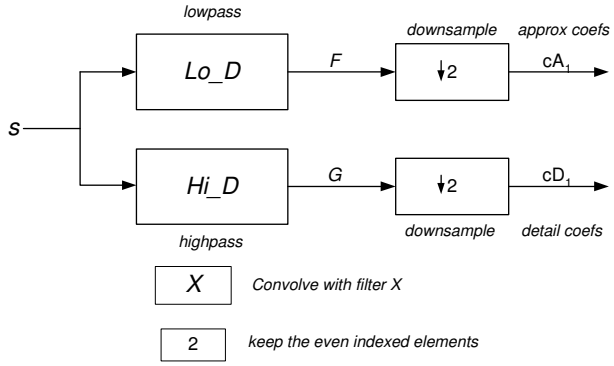


Figure 3.4: Wavelet decomposition

The downsampling of the signal components performed during the decomposition phase introduces a distortion called aliasing. It turns out that by carefully choosing filters for the decomposition and reconstruction phases that are closely related (but not identical), we can cancel out the effects of aliasing. These low and highpass decomposition filters along with their associated reconstruction filters are called quadrature mirror filters. Mathematically, let F_0 and F_1 be quadrature mirror filters (QMF). Then, for any finite energy signal $\mathbf{x}$,

$$||\mathbf{y_0}||^2 + ||\mathbf{y_1}||^2 = ||\mathbf{x}||^2 \tag{3.24}$$

where $\mathbf{y_0}$ is a decimated version of the signal $\mathbf{x}$ filtered with F_0 so $\mathbf{y_0}$ is defined by $\mathbf{x_0} = F_0(\mathbf{x})$ and $y_0(n) = x_0(2n)$, and similarly, $\mathbf{y_1}$ is defined by $\mathbf{x_1} = F_1(\mathbf{x})$ and $y_1(n) = x_1(2n)$. This property ensures a perfect reconstruction of the associated two-channel filter banks scheme (See [94])

The extension to the 2-D case is usually performed by applying a separable filterbank on the image. Typically a lowpass filter and a highpass filter are used. The convolution with the lowpass filter results in the approximation image and the convolution with highpass filter in a specific direction results in so-called details images. See Fig. 3.5 for explanation.

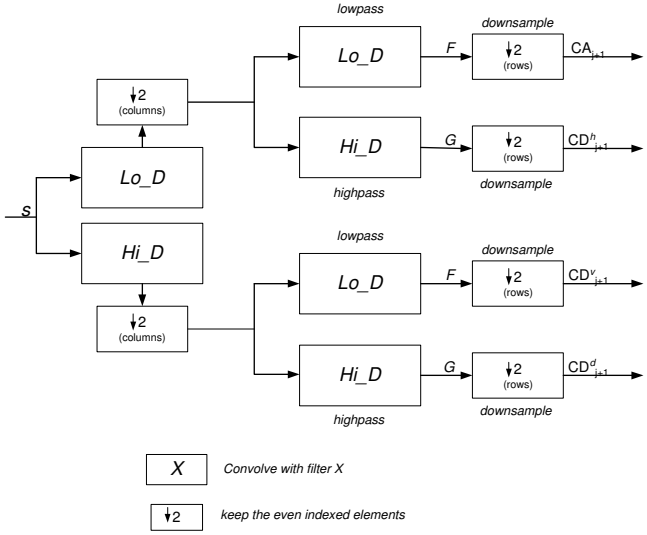


Figure 3.5: Two-dimensional wavelet transform

3.2.2 The Discrete Wavelet Decomposition

In the Wavelet decomposition, the image is split into an approximation and details images as shown in Fig. 3.6. The approximation image is then split itself into a second level approximation and details. For a n-level decomposition, the signal is

decomposed in the following way:

$$CA_n = [H_x * [H_y * CA_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{3.25}$$

$$CD_n^{(v)} = [H_x * [G_y * CA_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{3.26}$$

$$CD_n^{(h)} = [G_x * [H_y * CA_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{3.27}$$

$$CD_n^{(d)} = [H_x * [G_y * CA_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{3.28}$$

where $*$ denotes the convolution operator, $\downarrow 2,1$ the subsampling along the rows, and $\downarrow 1,2$ is the subsampling along columns. CA_0 is the original image. CA_n is obtained by lowpass filtering and is the approximation image at scale n. The details images $CD_n^{(h)}, CD_n^{(v)}, CD_n^{(d)}$ are obtained by bandpass filtering in a specific direction and thus contain directional detail information at scale n. The original image is thus represented by a set of subimages at several scales.

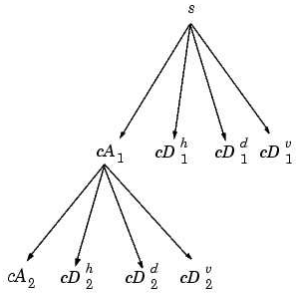


Figure 3.6: Two-dimensional, two-level wavelet decomposition tree

3.2.3 Generalized Wavelet Decomposition:Wavelet Packets

The wavelet packet approach is a generalization to the wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. In wavelet packet analysis, the details as well as the approximations can be split (the complete binary tree for one-dimensional three-level wavelet packet decomposition is shown in Fig. 3.7). Wavelet packets can be used to generate numerous expansions of a given signal. The most efficient representation of the given signal is then found according to an entropy-based criterion (other criteria have also been proposed). An example of such a two-dimensional wavelet packet tree decomposition is shown in Fig. 3.8. The naming convention of nodes in Figs. 3.7 and 3.8 is `(level, node number at that level)`. This is especially useful as the number of decomposition levels increases.

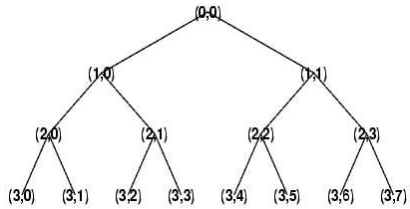


Figure 3.7: One-dimensional wavelet packet decomposition

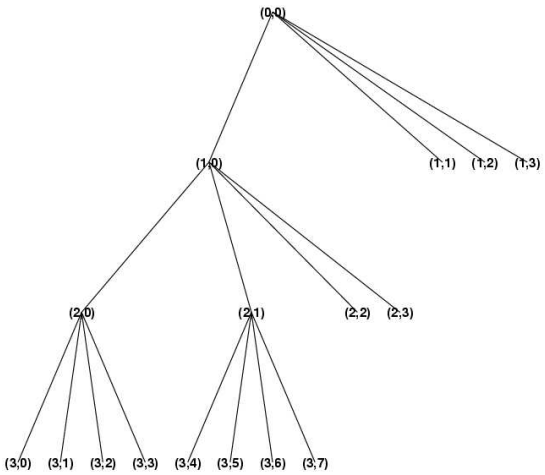


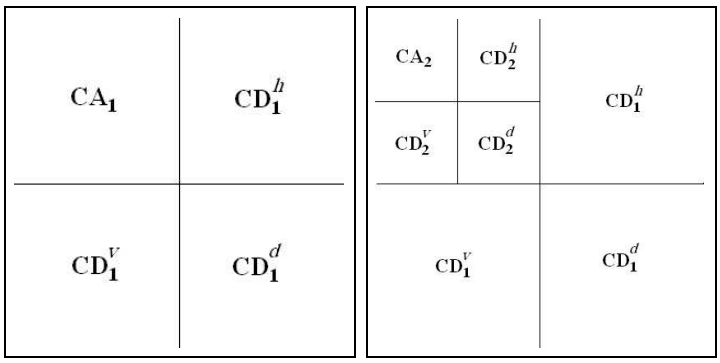
Figure 3.8: An example of a two-dimensional wavelet packet decomposition

3.2.4 Wavelet Decomposition for Images

The wavelet decomposed images in different subbands are usually represented in a peculiar manner. Fig. 3.9 shows a one-level and two-level wavelet decomposition of images. Fig. 3.12 shows an example of a two-level wavelet decomposition of image given in Fig. 3.11. The result of a full two-level wavelet packet decomposition of images is shown in Fig. 3.10. Fig. 3.13 shows a three-level wavelet packet representation of image given in Fig. 3.11.

3.3 The Multiscale Principal Component Analysis (MPCA)

All images are characterized by different frequency spectra; the information content is present in different frequency bands. The current systems that have been devel-



(a) One-level representation

(b) Two-level representation

Figure 3.9: Wavelet decomposed image representation

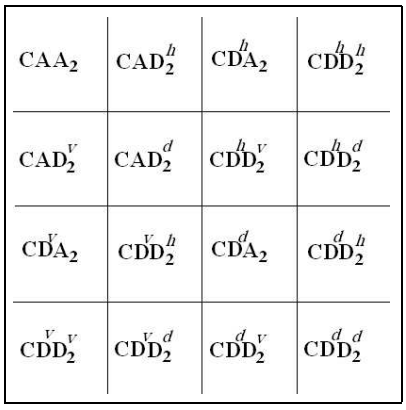


Figure 3.10: Two-dimensional wavelet packet image representation

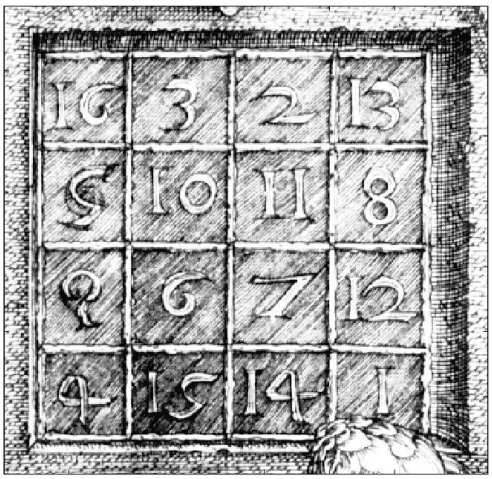


Figure 3.11: An example image

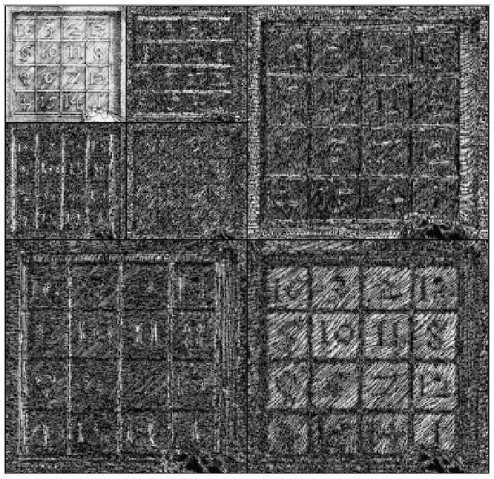


Figure 3.12: Two-level image wavelet decomposition example

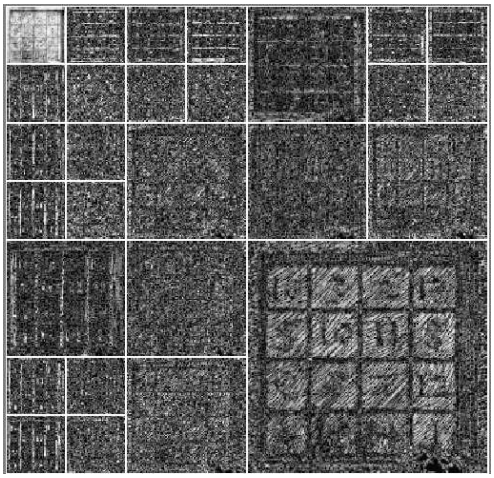


Figure 3.13: Three-level image wavelet packet decomposition example

oped work only on single scale, i.e., they do not utilize information from different frequency bands, instead, only the finest scale is used. This finest scale is defined by the sampling frequency of the data. Thus they miss the important information that can only be extracted by inspection at different scales.

Multiscale PCA (MPCA), as the name suggests, uses multiscale components of the input variables (data). That is, the system not only uses the original data at the finest scale but also considers it at different scales. For multiscale representation wavelets are very useful. For the MPCA, first the input data is represented into different scales using the Wavelet decomposition. The information at each scale is then sent as an input to a PCA system which de-correlates the data and provides a lower dimensional representation. Hence MPCA combines the ability of PCA to extract correlation or relationship between the variables (image pixels in our case), with that of orthonormal wavelets to separate deterministic features from stochastic

processes.

In pattern recognition systems, this representation can be used in a number of ways. As discussed earlier, the decomposed images can be used independently for training and recognition. The multiscale decomposition of images at each level gives one approximation image and three details images. We can either select the approximation images of all faces to train the system or use the detail images. The alternative approach is to find those scales that are important for recognition then combine these in an efficient manner to form an enhanced recognition system. The important scales are those that give the best recognition rates and hence play an important role in the overall recognition performance. The question is, obviously, how do we choose such scales. In chapter 6, we propose a new technique to obtain such selection.

3.4 Overview of Active Contour Models (Snakes)

Active Contour Models are also called snakes because they appear to slither across images. The Active Contour Models(ACM) belong to the general category of deformable models that use energy minimisation. From any starting point, the snake deforms into alignment with the nearest salient features of an image. These features in an image correspond to the local minima in the energy generated by processing the image [95, 96].

More precisely we can explain snake as a parametric contour($\mathbf{x}$) that deforms over a series of iterations. The $\mathbf{x}$ depends on two parameters

$$\mathbf{x}(s,t) \begin{cases} s = \text{space(curve) parameter} \\ t = \text{time(iteration) parameter} \end{cases}$$

The shape of the contour is influenced by the forces listed below:

- internal forces
- external forces and
- image forces

The total energy of the snake model at a particular iteration ($\mathbf{x}(s)$) is given by the sum of the energy for the individual snake elements [95]

$$E_{snake} = \int_0^1 E_{element}(\mathbf{x}(s))ds \tag{3.29}$$

The equation above can be written in terms of the three basic energy functionals:

$$E_{snake} = \int_0^1 E_{intern}(\mathbf{x}(s))ds + \int_0^1 E_{extern}(\mathbf{x}(s))ds + \int_0^1 E_{image}(\mathbf{x}(s))ds \tag{3.30}$$

The internal, external and image energy functionals are briefly explained below.

3.4.1 Internal Energy

The internal energy of a snake element is defined as

$$E_{intern}(\mathbf{x}) = \alpha(s)|\mathbf{x}_s(s)|^2 + \beta(s)|\mathbf{x}_{ss}(s)|^2 \tag{3.31}$$

This energy contains a first-order term controlled by $\alpha(s)$, and a second-order term controlled by $\beta(s)$. The first-order term makes the snake contract like an elastic band by introducing tension; the second-order term makes it resist bending by producing stiffness.

3.4.2 External Energy

In the absence of external forces, an active contour model simply collapses to a point like a strip of infinitely elastic material. We can use both manual and automatic supervision to control attraction and repulsion forces that drive active contour models to or from specified features. For example, a spring-like atractive force can be generated between a snake element and a point in an image using the following external energy term

$$E_{extern}(\mathbf{x}) = k|\mathbf{i} - \mathbf{x}|^2$$

In order to make part of an image repel active contour model we use

$$E_{extern}(\mathbf{x}) = \frac{k}{|\mathbf{i} - \mathbf{x}|^2}$$

3.4.3 Image Energy

The potential(image) energy P generated by processing an image $I(x, y)$ produces a force that can be used to drive snakes towards features of interest. Three different potential(image) energy functionals are described below. These attract snakes to

lines, edges and terminations.

$$P = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term}$$

Here w's are the weights for each of the energy terms.

The simplest potential energy that can be used to attract snakes towards lines is the image intensity $I(\mathbf{x})$.

$$E_{line} = \int_0^1 I(\mathbf{x})ds$$

The edges can be found by using a gradient-based energy functional as given below

$$E_{edge} = -\int_0^1 \left|\frac{\partial I}{\partial \mathbf{x}}\right| ds$$

The ends of line segments, and therefore corners, can be found using an energy term based on the curvature of lines in a slightly smoothed image $C(x,y) = G_\sigma(x,y) * I(x,y)$. If the gradient direction is given by $\theta = tan^{-1}(C_y/C_x)$ then the unit vectors along, and perpendicular to, the image gradient are given by:

$$\text{Tangent: } \mathbf{n} = \begin{pmatrix} cos\theta \\ \\ sin\theta \end{pmatrix}$$

$$\text{Normal: } \mathbf{n}_\perp = \begin{pmatrix} -sin\theta \\ \\ cos\theta \end{pmatrix}$$

The curvature of a contour in $C(x,y)$ can be written:

$$E_{term} = \int_0^1 \frac{\partial \theta}{\partial \mathbf{n}_\perp} ds$$

This energy formula provides a simple means for attracting snakes towards corners and terminations. Below (Fig. 3.14)we give an example showing a snake converging iteratively to the desired edges.

3.5 Summary

In this chapter, we discussed all of the mathematical techniques and tools that will be used in the thesis. A detailed mathematical explanation of the method of principal component analysis was first discussed, followed by a detailed explanation of Wavelet transform. In the end, the active contour models (snakes) technique for curve deformation was explained. In the next chapter, we will discuss a number of newly developed techniques for face and eyes detection.

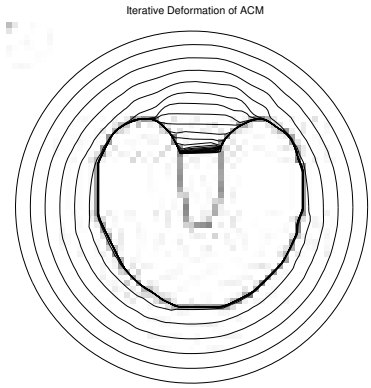


Figure 3.14: Iterative Deformation of ACM

Chapter 4

Proposed Face and Eyes Detection Techniques

4.1 Introduction

A robust face recognition system must be able to handle the different variations in the input face images. The input images can vary in size, orientation, etc. This brings in the importance of the process of normalization. Such a process normally involves the rotation of images to straighten them, and rescaling to a particular size. This can be done only when the faces have been detected from the input images and the eyes are localized. As such, before any normalization, the eyes need to be detected from the face image. The flowchart of Fig. 4.1 explains the process we are proposing to create the normalized images. Here it should be mentioned that

the process of face and eyes detection is generally dependent upon the type of input images. In order to achieve robustness, we used three different databases. In the next section, we will discuss some distinctive properties of these databases.

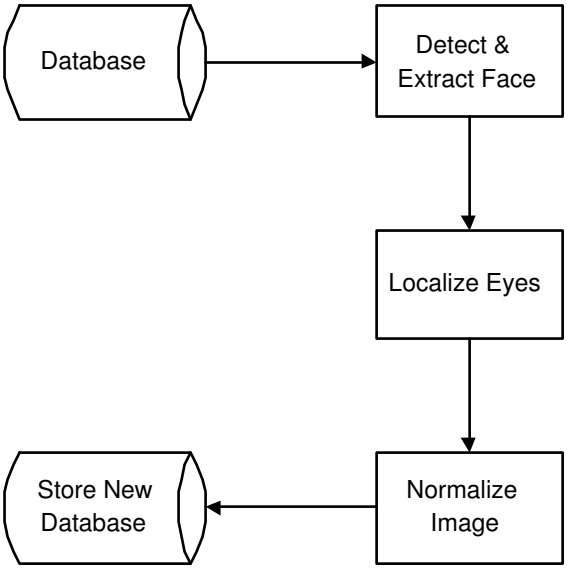


Figure 4.1: Flowchart showing the normalization process

4.2 Description of the Databases

We performed our experiments on three different databases. In Table 4.1, the names of these popular databases are listed.

Name	**Classes**	**Images per class**
ATT (ORL) [97, 98]	40	10
Yale [99]	15	11
AR [100]	126	30

Table 4.1: The face databases

In the 'ATT Database of Faces', (formerly 'The ORL Database of Faces') [97, 98],

we have images with medium-degree variations. By medium-degree variations we mean the images are with different expressions, slight rotations, and moderately varying illumination conditions. For some subjects, the images were taken at different times, varying the lighting and facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The ATT face database is different from the other two databases, Yale and AR, in the sense that the images contain only the face region. Since all the images in the database are of the same size, it is obvious that some of the original face images must have been *resized* to develop the database. It is apparent in this database that we do not need to detect the faces. Some sample images from this database are shown in Fig. 4.2. In contrast, the Yale and AR databases are comprised of face and shoulder images with a uniform background. The background only varies due to the different lighting effects. The AR Database [100] contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). The images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc., were imposed on the participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. The images in the Yale database

[99] were also taken with different facial expressions or configurations: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. Some sample images from these databases are shown in Fig. 4.3 and 4.4.

Figure 4.2: Sample images from the ATT face database

Due to the above dissimilarities, the face and eyes detection process for the ATT face database is not the same as for the other two databases. In the sections that follow, we will explain the eyes detection process for the ATT face database followed by the face and eyes detection for the Yale and AR databases. The normalization procedure is the same for all these databases and is discussed at the end.

4.3 The Proposed Face Detection Method

The Yale and AR databases are different from the ATT face database. The images in these databases consist of the face and shoulder regions. In order to perform face

Figure 4.3: Sample images from the Yale database

Figure 4.4: Sample images from the AR database

recognition, and other related tasks, we need only the face regions. Our intention was to build new database from these databases. The new database will consist of only the face regions. An example of the detected face region is shown in Fig. 4.5.

Figure 4.5: Example: detected face

The process that we developed for face detection is shown in the flowchart of Fig. 4.6, and the whole process is described in Alg. 4.1.

Face detection is performed by looking for the face boundary points and using such information to extract the face region from the whole image. The Active Contour Model (ACM) technique also known as the Snakes technique is used for face boundary extraction. This technique is applicable because of the near uniformity in the background of the images. The background in the Yale and AR databases is completely uniform, while in some images for the Yale database, it is affected by shadowing. The ACM technique can not be used if the image background is not

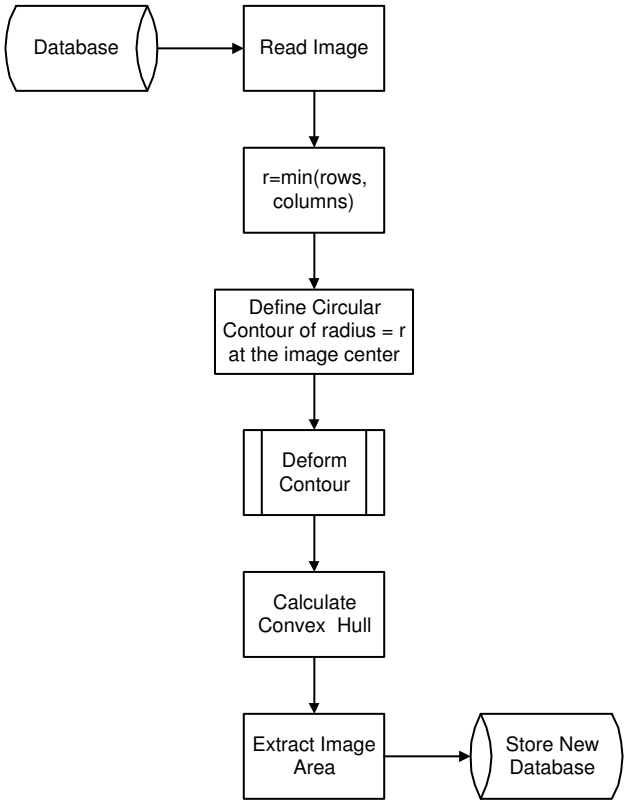


Figure 4.6: Flowchart: face detection

Algorithm 4.1 Face Detection

1: ReadImage I
2: $m = \text{rows}(I)$
3: $n = \text{columns}(I)$
4: Center $(cm, cn) = (m/2, n/2)$
5: $radius = \min(m, n)$
6: InitializeContour (x, y)= Circle$(radius, cm, cn)$
7: **for** $i = 1$ to k **do**
8: $\quad (x, y) = \text{DeformContour}(x, y)$
9: **end for**
10: $(x, y) = \text{ConvexHull}(x, y)$
11: ExtractImageArea $I(x, y)$

uniform. The process proposed for face detection is described below:

1. *Define an Active Contour Model around the face region.* This requires a rough estimate of the location of face in an image. This method is unapplicable if we do not, roughly, know the face position. This is the only disadvantage of the method. On the other hand, the ACM is able to detect faces in any orientation. For the Yale and AR databases, we know that the subject's head will be approximately at the center of image. Width and height of the image are first calculated and the minimum of both is then determined. Such value can then be used as the diameter of the initial ACM circle. Note that it is not necessary to start with a circle. We can define the ACM starting point with any shape. Since the human face is of elliptical shape, it is better to define the ACM with an ellipse. This will help converge the snake more accurately and more quickly around the face. We can select an elliptical shape ACM only when we know the orientation of face. Fig. 4.8 shows an elliptical shape defined by the ACM. Since we do not know the orientation of the faces a priori, we can not use the elliptical shape ACM. As such, the ACM in our case is initialized with a circle. An example is shown in Fig. 4.9.

2. *Start the snake convergence.* The algorithm then searchs for the nearest edges to fit upon. As discussed in Sec. 3.4, the shape of the ACM is guided by the potential energy calculated using the image intensity values. This helps the

curve fitting around the face boundary. Since the process is iterative, it takes the snake a number of iterations to deform and fit on the face boundary. Fig. 4.10 shows the ACM after convergence.

3. *Extract the face region defined by the calculated boundary.* It was observed that in some cases the boundary is not smooth. There are some concavities in the detected face boundary. (See for example Fig. 4.7). This results in the loss of some of the parts of the face. Our experiments show that this is mainly due to the variations in the illumination. In some cases, the high intensity light is reflected from the face making it similar to the white background. The boundary of the face disappears and results in a false face boundary. The solution we propose to this problem is to determine the convex hull for the face boundary. A convex hull of an image region is the smallest convex set containing the image. Some examples of the detected face boundaries and their corresponding convex hull are shown in Fig. 4.11.

4. *Form an image from the extracted face region.* The convex hull of the detected face boundary is used to define a new image consisting of only the face region. The dimensions of this new image are selected to enclose the whole face defined by the convex hull. It should be noted that the ACM technique can detect faces in any orientation. Some examples for the detection of rotated faces are shown in Fig. 4.11.

Figure 4.7: Example image boundary detection with concavities

Figure 4.8: Active Contour Model when initialized as an ellipse

Figure 4.9: Active Contour Model when initialized as a circle

Figure 4.10: Final position of the ACM

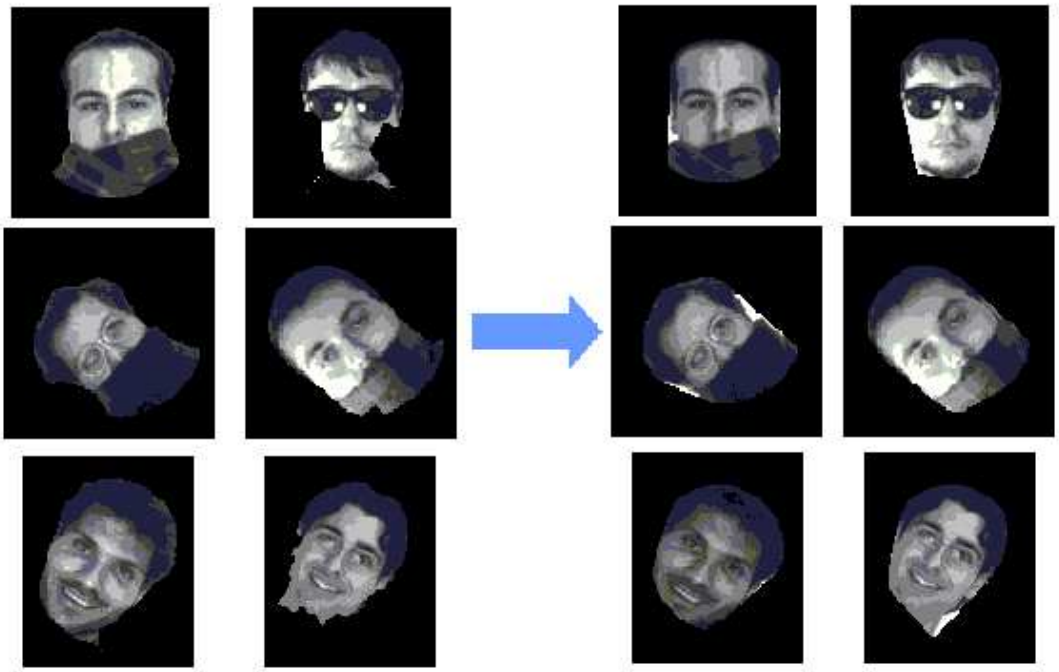


Figure 4.11: Detected face boundaries and their convex hulls

Once the ACM is complete, the next stage of eyes detection from the face images is initiated. The proposed methods for the detection of the eyes are discussed in the next section.

Experimental Results

The AR database consists of the face and shoulder images. We used 600 images from the AR database (310 for men and 290 for women). A face was assumed to be detected accurately only if the whole face area was captured within the boundary. This consists of the main facial features like eyes, mouth, and nose. The face detection accuracy that we achieved using the proposed technique was quite satis-

factory reaching **98.33**%. In a limited number of cases, the boundary of the faces was not well-detected because of the varying illuminations. At some points the face boundary disappeared because of the same shade of background and skin.

4.4 The Proposed Eyes Detection Methods

4.4.1 Method 1: Template Matching

Two techniques have been proposed for detecting the eyes from face images: the first one is based on template matching and the second is based on wavelets. Faces with on-the-plane rotation can easily be detected with the technique discussed in the previous section. Eyes detection algorithms from such images are expected to work with rotated images. The requirement, obviously, complicates the process of eyes detection. The eyes detection technique proposed in this section has been developed to work with face images, however with minor change it can accomodate faces only images such as the case of the ATT database. The proposed algorithm for eyes detection is outlined in Alg. 4.2 and explained below.

It is very rare to get a face image upside down in real life. Keeping this fact in mind, we put a restriction on our algorithm to work only with images that are rotated within a range of 180° i.e. from 90° clockwise to 90° counterclockwise as shown in Fig. 4.12. To simplify the process, we first rotate the input image to make it approximately straight.

Algorithm 4.2 Eyes Detection

1: Input detected face image I
2: $m' = \text{rows}(I)$
3: $n' = \text{columns}(I)$
4: Generate binary mask M from I
5: Find Principal Components (v_1, v_2) of M
6: Rotate I to make the major PC, v_1 horizontal
7: $I' =$ Top 60% of I
8: Divide I' into two parts I'_L and I'_R
9: Load L and R, the left and right eye templates
10: Resize Templates

$$m'_T = m_T \times \frac{m'}{m}$$
$$n'_T = n_T \times \frac{n'}{n}$$

11: Calculate correlations C_L between L and I'_L, and C_R between R and I'_R
12: Left eye center = Location where C_L is maximum
13: Right eye center = Location where C_R is maximum

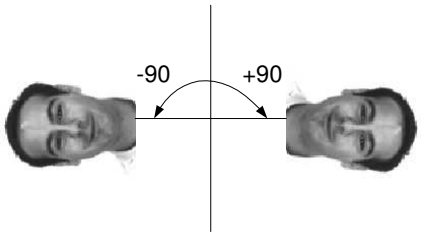


Figure 4.12: Range of possible rotations

For every input face image, we create a binary mask. This mask is created by setting all points within the detected face boundary to white and the other pixels to black. An example is shown in Fig. 4.13. This mask is of elliptical shape as human faces are elliptic.

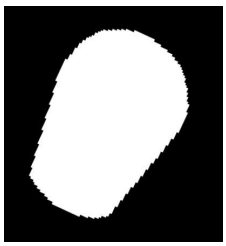


Figure 4.13: Binary mask for pose estimation

From this mask image, we can easily predict the direction of the first two principal components. The principal components for a typical face mask are shown in Fig. 4.14. The minor axis is approximately in the direction of eyes while the major axis is perpendicular to it. We can use this information to rotate the original image such that it becomes straight. Once the image is straightened, it is easy to detect the eyes.

Two separate templates are used representing the left and right eyes. These templates are shown in Fig. 4.15. These templates were obtained from the test images and can be used with any database. The search area is significantly reduced by selecting the upper 60% area of the face and further dividing it into two parts - left and right. As such, the search can be carried out in parallel for both the left and right eyes. Here, we have assumed that the size of the face images is m rows $\times$

n columns. If the input face is of different size, then we resize the templates. For example, if the input face image is of size $m' \times n'$, the eye templates are resized to

$$m'_T = m_T \times \frac{m'}{m}$$
$$n'_T = n_T \times \frac{n'}{n}$$

where m_T and n_T are the dimensions of the eye templates.

The correlations between each of the eye templates and the corresponding upper face region are calculated. The region giving the highest correlation is declared as the eye region. Some examples of detected eyes are shown in Fig. 4.16. Note that this technique is not very accurate with subjects wearing glasses.

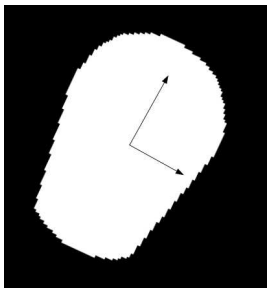


Figure 4.14: Principal components of the binary image

Figure 4.15: Left and right eyes templates (20×35) pixels

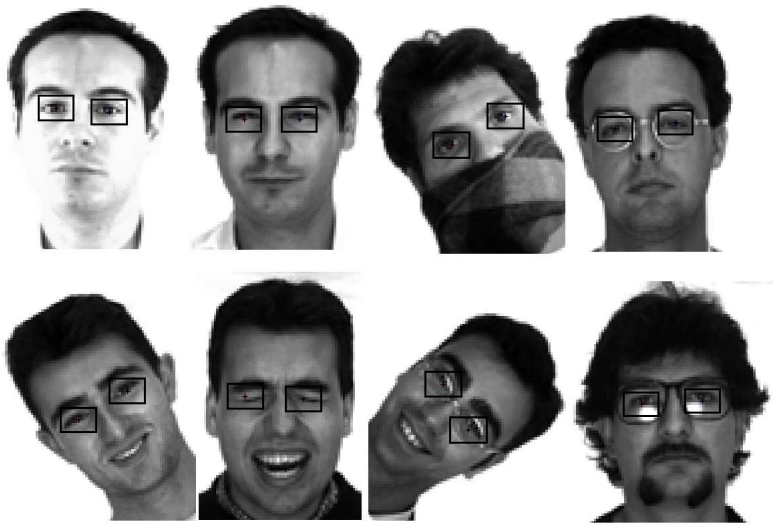


Figure 4.16: Examples: deteced eyes

Experimental Results

For the 600 images (1200 eyes) from the AR database, the detection accuracy was about **87.33**%. Fig. 4.17 shows the histogram for the error in distances of detected eyes center. This error is simply the differnce between actual eye center and the detected eye center. The threshold for correct detection of eyes center considerd here is 15 pixels. If the distance between the detected eye center and the original eye center is greater than this threshold, we do not consider it an eye center.

4.4.2 Method 2: Wavelet-based Eyes Detection

As stated earlier, we do not need face detection when using the ATT database. As such, we discuss here a technique that detects directly the eyes. The flowchart in Fig. 4.18 outlines the overall process for the ATT database.

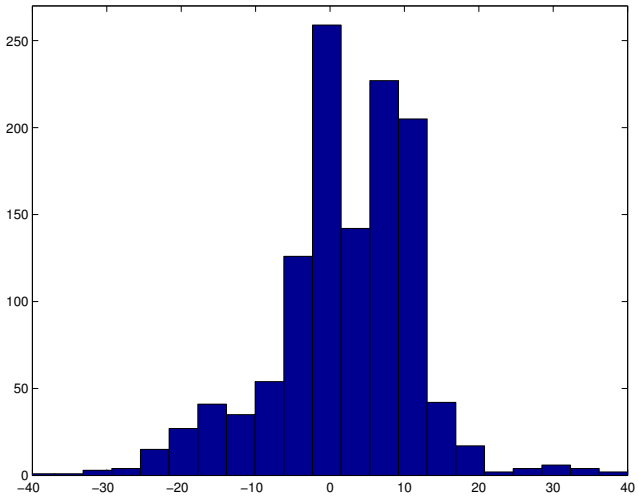


Figure 4.17: Distribution of error in detected eyes centers

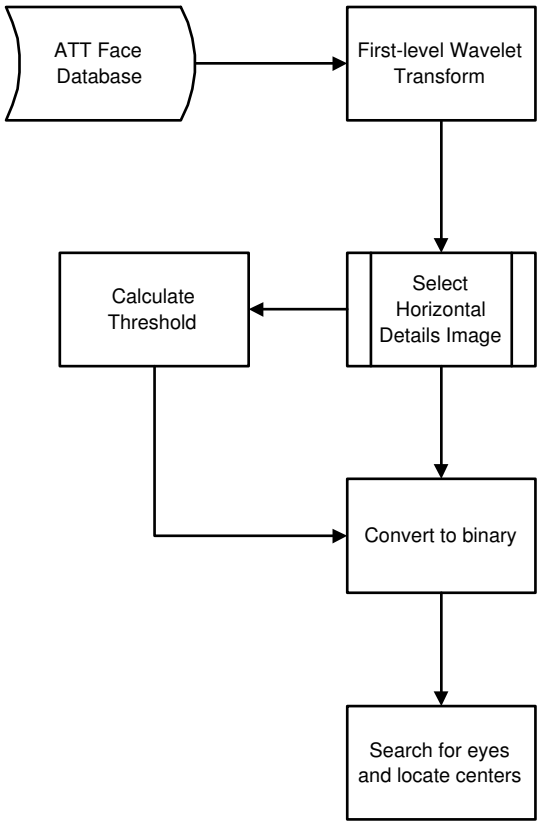


Figure 4.18: Flowchart for eyes detection process (ATT)

We propose here to detect the facial features by first converting the input image into a binary image. A threshold is needed for such conversion. We are looking for an appropriate threshold that will consider all prominent features in the input image and highlight these in the binary image. A good binary image representation can be obtained if the features that need to be detected are of high contrast in the input image. As such we need first to transform the input image into a high contrast image. This can be performed using the Wavelet decomposition. Fig. 4.19 shows a one-level wavelet decomposition of an input image. The facial features are of high contrast in the horizontal detail image. As such, we can conclude that the horizontal detail image is a good candidate as binary representation of the original image. The filters used here for the wavelet decomposition are found to be very effective in facial feature detection. These filters, proposed by [101], are given below:

$$\begin{aligned} H(z) = {} & 0.853 + 0.377(z + z^{-1}) - 0.111(z^2 + z^{-2}) \\ & - 0.024(z^3 + z^{-3}) + 0.038(z^4 + z^{-4}) \end{aligned} \tag{4.1}$$

$$G(z) = -z^{-1}H(z^{-1}) \tag{4.2}$$

where $H(z)$ and $G(z)$ are the lowpass and highpass filters, respectively. The horizontal details image are obtained when we apply high-pass filter in the vertical direction, and a low-pass filter in the horizontal direction. The reason for selecting this particular detail image is obvious as most of the facial features are in horizontal direction. First level decomposition gives a good detail image with prominent edges

that are useful specially for the facial feature extraction. The proposed threshold is obtained as the average of the highest and the lowest gray level of the detail image. The resulting binary image, shown in Fig. 4.20, consists of only high intensity points mainly showing the boundaries of eyes, mouth, nose etc. Some other edges are also visible depending upon the quality of the original image. A morphological opening operation is performed to remove these points at the border and some isolated points to get the final binary image as shown in Fig. 4.20.

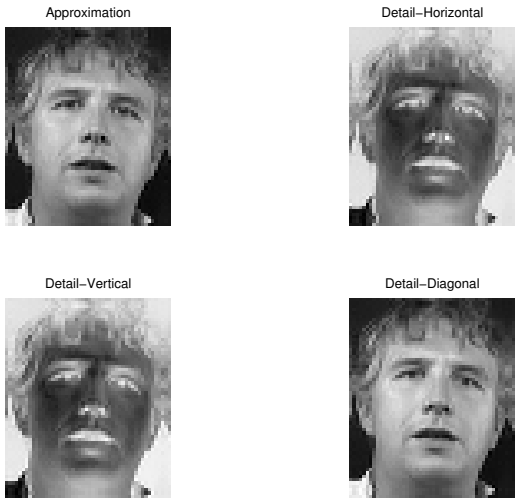


Figure 4.19: One-level wavelet decomposition

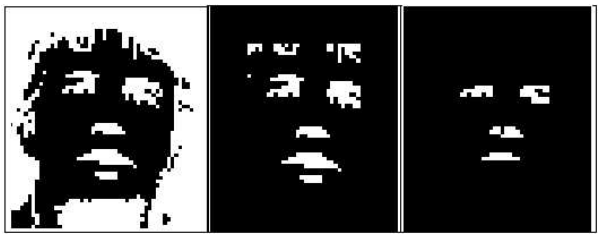


Figure 4.20: Binary Image after applying threshold, removing border points, and morphological operations

It is obvious that the position of the eyes can now be determined if we search for non-black points both row-wise and column-wise. This way, we separate both eyes regions and estimate the centers by calculating the centroid of these eyes regions. Once the centers of eyes are known in the binary image, we can map the centers in the original image as well. Some results are shown in Fig. 4.21. This completes the eyes detection stage for the ATT database.

Figure 4.21: Successful detection of eyes

Experimental Results

For the 400 images (800 eyes) from the ATT database the detection accuracy achieved is **85.75**%. The threshold for correct detection of eyes center that we considerd is same as discussed earlier in section 4.4.1 (15 pixels). We consider a detected eye center as a success if the distance between it and the actual center is smaller than such a threshold.

	Template Matching	Wavelet-based
AR database (1200 eyes)	87.33%	92.33%
ATT database (800 eyes)	82%	85.75%

Table 4.2: Summary of Eyes Detection Results

4.5 The Problems of Pose and Normalization

In our work, we have tackled the pose problem only for on-the-plane rotations. Experimental results suggest that even a slightly rotated face image degrades system performance tremendously when given as an input. This make it very important to remove any rotation effects from the input image. An algorithm for solving the problem of on-the-plane pose rotation is proposed below.

Algorithm 4.3 Pose Invariance

1: Find eyes centers $L(m_L, n_L)$ and $R(m_R, n_R)$.
2: Define line segment T between centers L and R.
3: Find angle, θ, between T and horizontal.
4: **if** $\theta \leq 90^\circ$ **then**
5: rotate image anticlockwise by θ degrees
6: **else**
7: rotate image clockwise by θ degrees
8: **end if**

Once we have the centers of eyes we normalize each image such that the line between the two centers is horizontal, as shown in Fig. 4.22. The rotation is performed in clockwise or counterclockwise manner depending upon the alignment of the eye centers. The rotation is performed with respect to the center of the image. Interpolation is needed for calculation of the resulting rotated image pixel values. Here, we have used the nearest-neighbour interpolation method.

4.6 Creation of the Eyes Database

Let the distance between centers of eyes be d. After rotating the whole image, we use this distance (d) to extract a "useful area". Based on extensive experimental results, we found that the optimal size of the eyes region should be $1.8d$ wide by $0.6d$ high. The details are shown in Fig. 4.23. The figure shows that, using these distances, we are able to cover very well the area containing eyes and the eyebrows. This is done with all images which are then resized to standard 30×75 images. Some samples from the resulting database are shown in Fig. 4.24.

Figure 4.22: Rotating the image

4.7 Summary

In this chapter, a face detection technique based on the ACM is presented followed by two different techniques for eyes detection. These include template-matching and a wavelet-based technique. Wavelet-based technique performed better than the template-matching. The eyes centers are very useful in performing image normalization. After presenting the two eyes-detection techniques, image normalization is

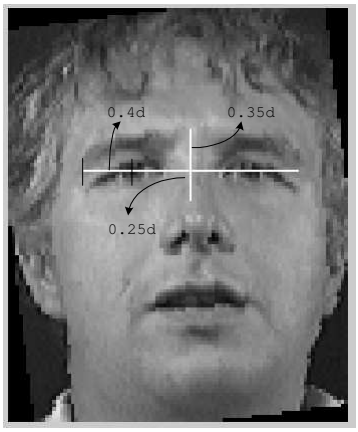


Figure 4.23: Extraction of eyes area

Figure 4.24: Eyes database samples

discussed. In the end, the process for creating an eyes database is explained and some database samples are shown. Such database will be made available to the research community worldwide.

Chapter 5

Face and Eyes Recognition using PCA

5.1 Introduction

In this chapter, the classical eigenface technique is used for face recognition. Face recognition is performed to evaluate the performance of our proposed pose invariant and illumination invariant techniques. In addition, the eigeneyes technique-a variant of eigenfaces which uses eyes as input images, is also implemented to compare the face recognition results with those of the eyes recognition. In the end, the method of least-squares is proposed as an improvement to the classical eigenfaces technique.

5.2 System Description-Face Recognition using Eigenfeatures

This process involves the use of principal component analysis (PCA). We divide the face images database into two sets, a training set and a testing set. The training set is comprised of a particular number of images from each class. The remaining images are used for testing. Before training, we compute a mean image from these training images and subtract it from all images. An example mean image for the ATT face database is shown in Fig. 5.1. The mean is removed in order to simplify the implementation of PCA. In the process of PCA, we first convert the two dimensional images into vectors by concatinating the rows. Let $\mathbf{x}_i$ denote the ith image in vector form of dimension $N \times 1$, then the estimated mean image (for a database of M images), shown in Fig. 5.1, is calculated as:

$$\mathbf{m}_\mathbf{x} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i \tag{5.1}$$

Let $\mathbf{e}_i$ and λ_i be, respectively the eigenvector and corresponding eigenvalue of $\mathbf{C}_\mathbf{x}$-the estimated covariance matrix given by:

$$\mathbf{C}_\mathbf{x} = \frac{1}{M} \left[\sum_{i=1}^{M} \mathbf{x}_i \mathbf{x}_i' \right] - \mathbf{m}_\mathbf{x} \mathbf{m}_\mathbf{x}' \tag{5.2}$$

Then a transformation matrix $\mathbf{A}$, whose rows are the eigenvectors of $\mathbf{C}_\mathbf{x}$ can be used to project an image into the eigenspace formed by the eigenvectors $\mathbf{e}_i$'s. This transformation is $\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_\mathbf{x})$. In this transformation, dimension reduction of

images is performed while preserving the components that contribute most to the variance in the original data. We can select an arbitrary number of significant eigenvectors for projecting an image into the eigenspace. As the number of eigenvectors increases, a better representation of the image is achieved. Usually the number of images used for training (M) is less than the number of pixels (N) in an image. In order to simplify the expensive process of eigenvectors calculation dimensionality reduction of $\mathbf{C_x}$ matrix is performed. The overall recognition rate for the PCA decomposition is usually displayed as a function of the number of eigenvectors. It can be seen easily that the recognition rate improves as we increase the number of eigenvectors to form eigenspace (see Figs. 5.5, 5.7). We call the eigenvectors of faces the eigenfaces and those of eyes the eigeneyes.

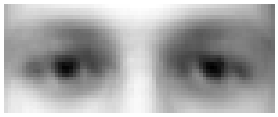


Figure 5.1: Mean eye image for the ATT eyes database

In our work, the Eigeneyes are found by applying PCA on the eyes database. This is what creates the eigenspace. The process of finding the eigeneyes is the training stage of system. Once the system is trained, we test it on a number of images. Obviously, these test images are not used in the training process. These test images (after removing mean) are projected onto the eigenspace. Similarly all training images in the database are projected onto the eigenspace. We then compute the distance between the projection of the unknown (test) image and all

the projections of the training images from database. The class which gives the minimum distance is declared as the class of the unknown test image. Euclidean distance metric is used in this case. The first four eigenfaces and eigeneyes can be seen in Fig. 5.3 and 5.2, respectively. The whole process is described in schematic form in Fig. 5.4.

Figure 5.2: The first four eigenfaces (Yale)

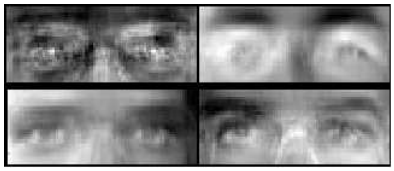


Figure 5.3: The first four eigeneyes (ATT)

The main goal of this chapter is to analyze the performance of PCA using whole faces and using eyes alone. Different configurations and setups will be discussed in addition to a discussion of pose invariance. Please note that all experiments of face and eyes recognition were performed with a set of four training and six testing images.

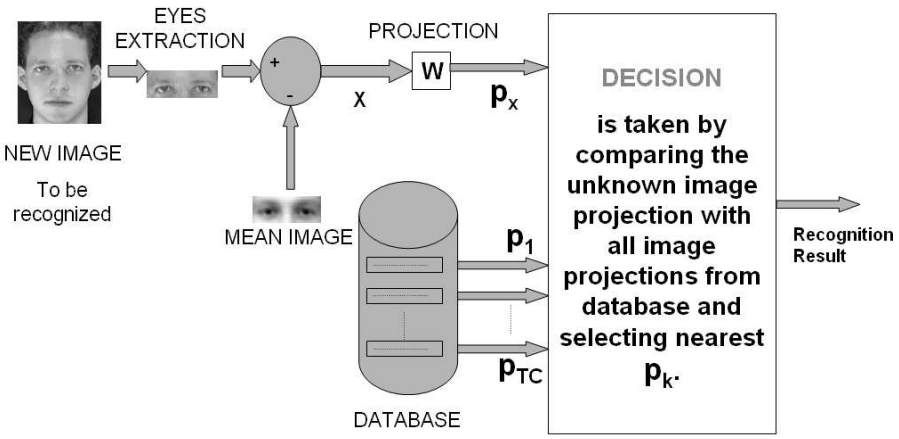


Figure 5.4: Schematic diagram (eyes recognition using PCA)

5.3 Recognition using Face Images

In the experiments that follow we will be using face images for recognition purposes. The experiment with eyes will be discussed in the next section. The result given in Fig. 5.5 shows the recognition rate as a function of the number of eigenvectors. The recognition rate saturates after some specific number of eigenvectors. This is because most of the energy is contained in the first few eigenvectors. Fig. 5.6 shows an example of the calculated eigenvectors energy for the ATT database.

The results for the ATT database are shown in Fig. 5.5. These correspond to a system that was trained using four images per subject. For testing, six images per subject were used. The four images selected for the training purpose were fixed for all classes. Images in all classes are arranged in a particular manner. This means that the set of conditions for an image at a particular position, in a class, will be the same for all images at that position in other classes. So for the result shown

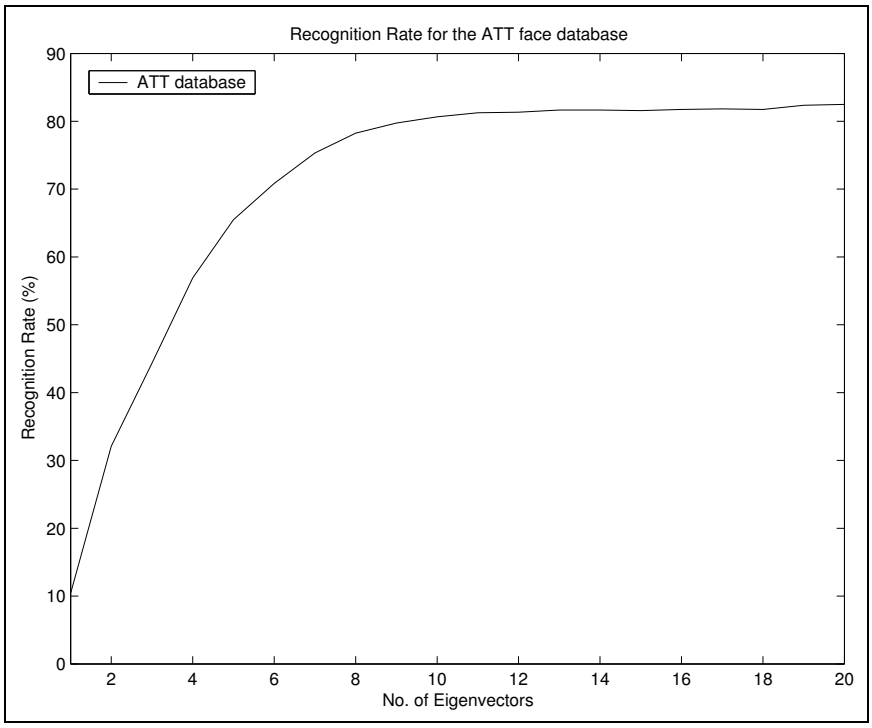


Figure 5.5: Recognition rate, simple PCA (ATT database)

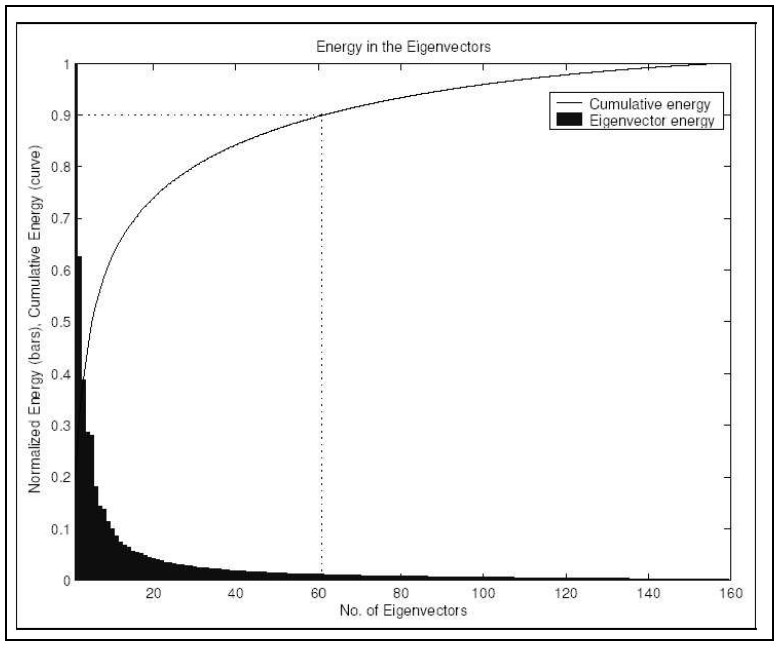


Figure 5.6: Eigenvectors energy details

above, the set of conditions for the four images selected for training are same.

More representations were also obtained using the *leave-one-out* technique to increase the number of datasets. This increases the statistical significance of the results. In this technique, we select different combinations of the training and testing sets. For example, in the ATT database there are 10 images per class. If we select, four images per class for training then there will be $\frac{10!}{4!(10-4)!} = 210$ possible number of sets. As an example some of the possible sets are given in Table 5.1.

Training Sets	**Testing Sets**
1, 2, 3, 4	5, 6, 7, 8, 9, 10
2, 3, 4, 5	1, 6, 7, 8, 9, 10
3, 4, 5, 6	1, 2, 7, 8, 9, 10
4, 5, 6, 7	1, 2, 3, 8, 9, 10
5, 6, 7, 8	1, 2, 3, 4, 9, 10

Table 5.1: Examples of different combinations of images for training and testing

The result shown in Fig. 5.7 is obtained using 20 different training sets on the basis of the leave-one-out technique. This result is statistically more meaningful then that of Fig. 5.5.

The importance of this technique is highlighted by this result. We can see that by using the first four images for training we got a maximum recognition accuracy of about 82%. When the leave-one-out technique was used the recognition accuracy shoots toward 90%. This shows that it is wrong to consider only one dataset for training any system. To get a better and more realistic view of a system, the leave-one-out technique should be used.

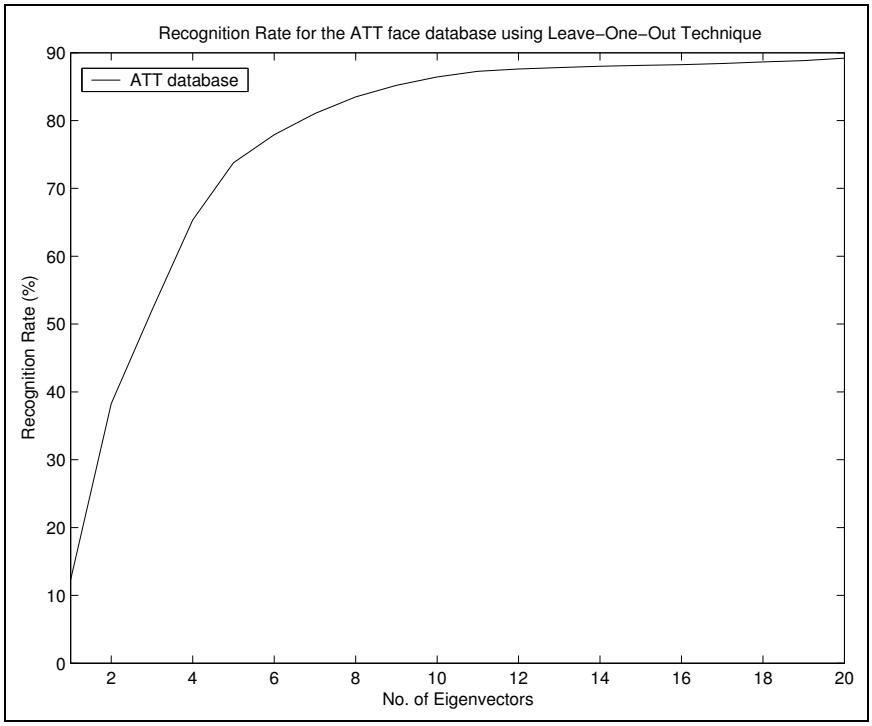


Figure 5.7: Recognition rate, leave-one-out PCA (ATT database)

5.3.1 Pose Invariance

The most common and serious problem faced by any face recognition system is that of pose variations. Sometimes the face images input to a face recognition system are tilted or rotated. By this we mean that the face in the input image is not straight. While in some cases the size of faces in input images is different. These variations affect badly the performance of face recognition systems. The PCA-based system discussed in the last section is also prone to such changes. We therefore propose techniques to handle these variations. In what follows, we analyze the performance of a PCA-based system for such conditions and propose our techniques to counter these problems.

Effect of Rotation

We will now discuss the effects of different orientations of face images on the recognition accuracy. We first explain the experimental setup. The training set is comprised of straightened frontal face images while the test set is comprised of rotated face images. In total, 13 test image sets were considered in these experiments. Fig. 5.9 shows that the original image (shown in Fig. 5.8) is rotated in 13 different degrees. These range from $-60°$ to $60°$ with steps of $10°$.

Figure 5.8: Original frontal face image

Figure 5.9: Twelve rotated versions of the frontal face image

It is obvious that as the rotation angle increases, the recognition accuracy drops. The smallest value of recognition rate occurs for greatest rotation. We see now the need for our proposed algorithm. We can improve the performance of PCA-based

system significantly if the input rotated face images are converted to frontal face images using our techniques discussed in Sec. 4.4.1 and 4.5.

To check the performance and accuracy of PCA-based system we performed experiments by supplying these rotated images to the system. The recognition rate results are shown in Fig. 5.10. As expected, the recognition rate drops sharply with the increase in angle difference. For sake of clarity results are shown only for angles $-10°$ to $-60°$. The results for the other angles can be inferred from the graph. They will be similar to their counterparts with oppositie signs. This shows how bad is the affect of even a small change in the angle on the performance of PCA-based system. The PCA-based system completely fails when rotated images are provided as input. This highlights the need for our proposed algorithms in Sec. 4.4 and 4.5.

Effect of Scale

The next important property that can affect the recognition accuracy is image scale. In real systems we can have input images with different scales and resolutions. It can be deduced that images with low scale will result in degraded performance. A number of experiments were performed to study the effect of different scales on recognition.

The system was trained with original scale frontal face images. Testing was performed with scaled images. Five new databases of scaled images were created from the original ATT frontal face database for the testing purpose. The images

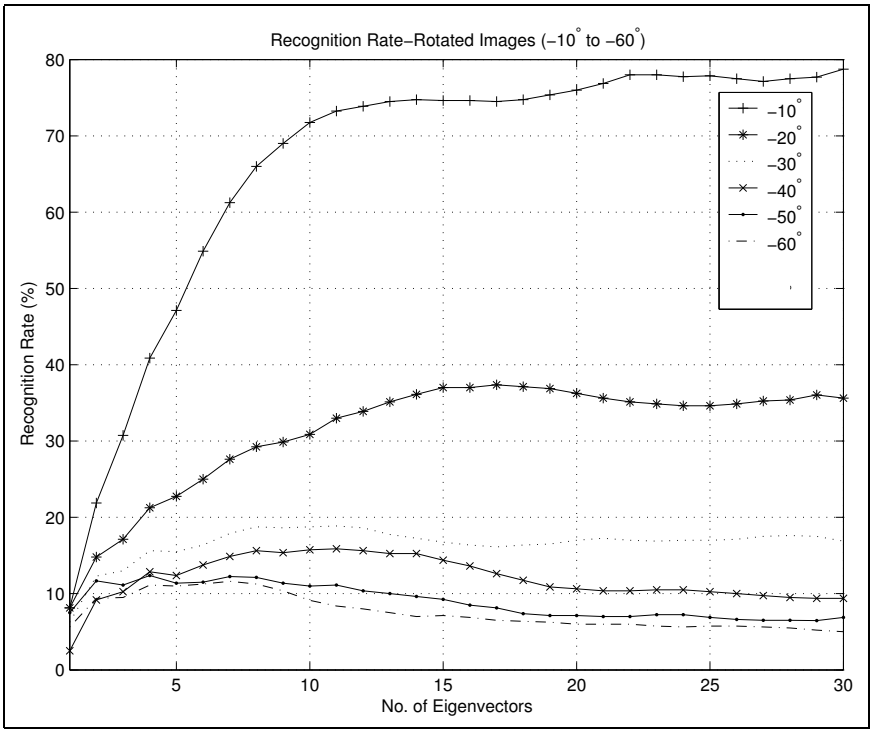


Figure 5.10: Recognition rate results for rotated images from $-60°$ to $20°$ (ATT database), without pre-processing.

were scaled down to 50% with steps of 10%. Fig. 5.11 shows same face image at five different scales.

Figure 5.11: ATT face image example with five different scales

It should be noted here that the input images to our system must be of same size as training images. This requires that we resize the scaled images before processing them for recognition. The resized versions of the images are also shown in Fig. 5.12. The degradation in images is easily visible for higher downscaling.

Figure 5.12: ATT face image example with five different scales, resized to same size (scale decreasing from left to right, top to bottom)

The recognition results are shown in Fig. 5.13. We observe that there is not much difference in the recognition accuracy with change in scale. Note that the plot limits are decreased to have a closer view of the recognition rate values. This result is obtained when six images were used for training and four images for testing.

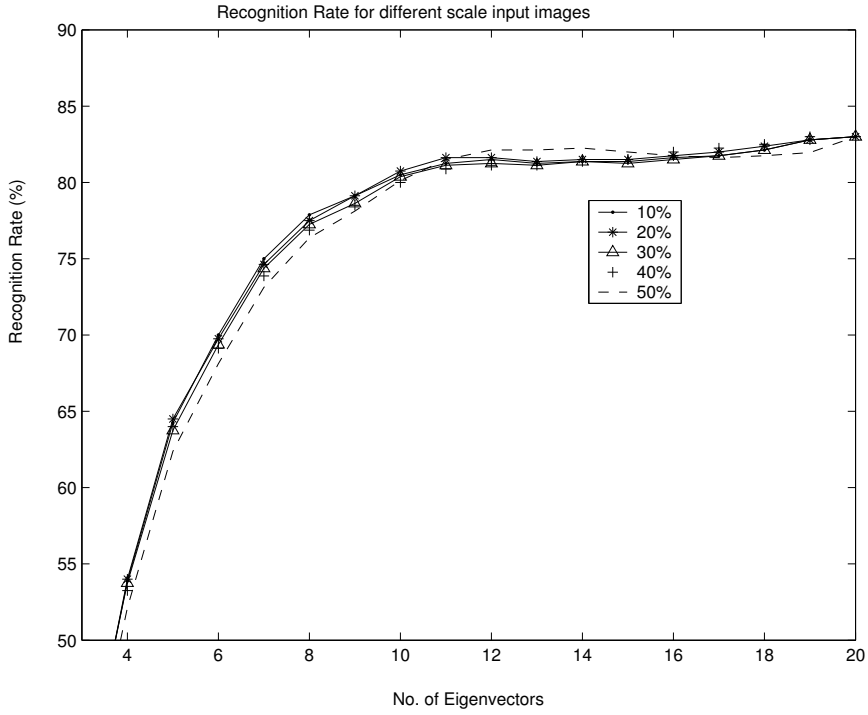


Figure 5.13: Recognition rate for scaled images (ATT database)

5.4 Recognition using Eyes Images

Many techniques have been developed that vary in terms of complexity and efficiency for face recognition [102]. Most of the face recognition systems have a drawback of long training and recognition times. Some face recognition techniques developed use the complete face image [45] while others perform operations on local features like eyes, nose, mouth etc [43, 42]. It has been observed experimentally that good recognition rate can also be achieved using only local facial features. Face features change over time. There are also other possibilities like facial hairs and beard that can affect the recognition process. Unlike the complete face, eyes do not change significantly with time. So it is a good idea to use eyes for face recognition. It helps in reducing complexity and improving efficiency (smaller data size).

The experiments that were performed for faces were also performed for the eyes database. This eyes database was created from the ATT face database using our technique discussed in Sec. 4.6. The results show that using only the eyes, we are able to get comparable recognition rate. The reduction in recognition accuracy is much less than the reduction in image size. We are using only 22% of the original image size, and are still able to get a recognition rate of 75.13%. This reduction in size resulted in a much less computational complexity of our system. We were able to reduce the computational complexity of PCA-based system by 5.45 times. Fig. 5.16 shows the computational complexity as a function of the number of image

pixels i.e. $NlogN$, where N is the number of image pixels.

Our proposed LS-PCA technique (discussed in the next section), when applied to the eyes also outperforms the simple PCA technique, giving 81.25% recognition accuracy. In fact we are able to get better results than the best ever reported result in literature for eyes by [55]. The results for the simple PCA are shown in Fig. 5.14 while those for LS-PCA are shown in Fig. 5.15. In Fig. 5.21 the result for LS-PCA using eyes is compared with the traditional PCA results using both face and eyes.

5.5 The Proposed Method of Least-Squares PCA

5.5.1 Introduction

In practice, a number of training images are given for each class (person). PCA does not take into account this information as it assumes that all images represent different persons. To take into account class information, we propose a new method based on the least squares estimate of the unknown image projection to improve classification. In this method, the process requires a model that relates the unknown image to the images in each class. This can be done with the help of one or more coefficients. The image obtained by using these coefficients is called the fitted image. To avoid the complexity that occurs due to the large size of images, we use the projections of the images onto the eigenspace. The difference between the projection of the unknown image vector ($\mathbf{x}$) of dimension $K \times 1$ and the projection of the fitted

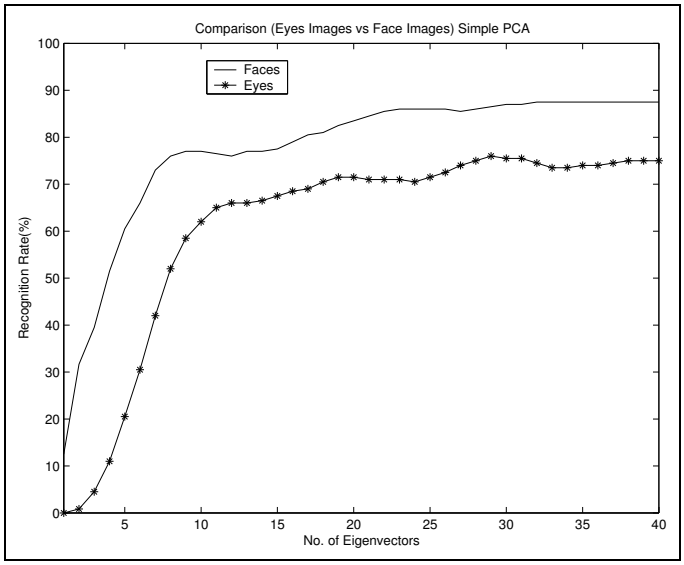


Figure 5.14: Eyes vs faces (PCA) (ATT database)

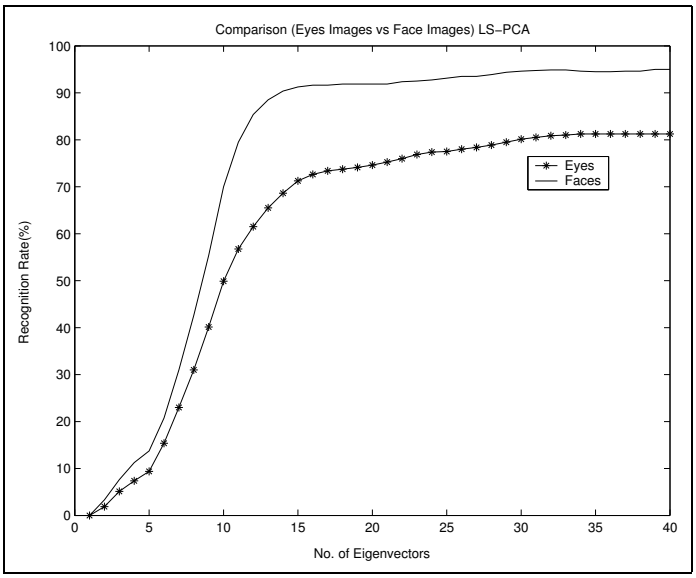


Figure 5.15: Eyes vs faces (LS-PCA) (ATT database)

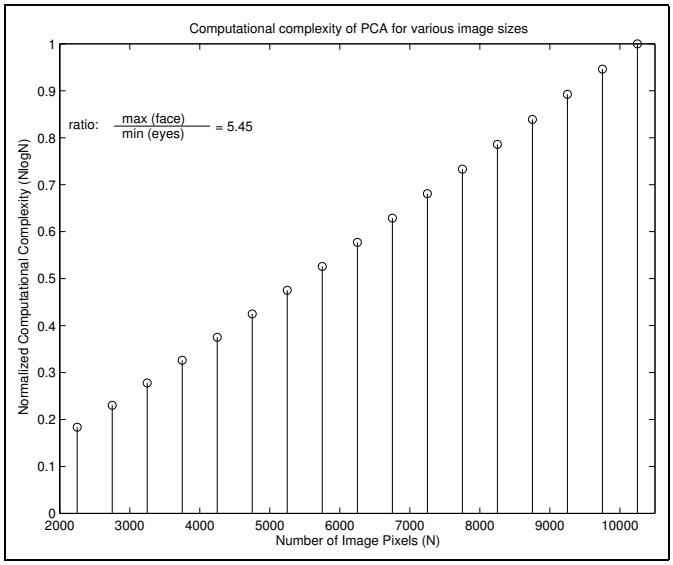


Figure 5.16: Reduction in computational complexity when using eyes

image vector ($\hat{\mathbf{x}}$) is defined as error (ϵ), also called residual.

$$\epsilon = \mathbf{x} - \hat{\mathbf{x}} \tag{5.3}$$

Note that $\mathbf{x}$ and $\hat{\mathbf{x}}$ are in vector form. The least squares method minimizes the summed square of residuals to obtain the coefficient estimates.

5.5.2 Analytical Formulation

Lets say we decide on using k training images per class, then we need k coefficients $(\alpha_1, \alpha_2, ..., \alpha_k)$. These can be found using the following equation:

$$\mathbf{x} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + ... + \alpha_k\mathbf{x}_k + \epsilon \tag{5.4}$$

$$\mathbf{x} = [\mathbf{x}_1\mathbf{x}_2...\mathbf{x}_k] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} + \epsilon$$

$$\mathbf{x} = \mathbf{X_c}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

Here $\mathbf{X_c}$ is a matrix, unique for each class, formed by the projections of k images for that class and $\boldsymbol{\epsilon}$ is the error vector. The LS estimate of the coefficient vector, $\hat{\boldsymbol{\alpha}}$, can be found by:

$$\mathbf{X_c}^T\mathbf{X_c}\hat{\boldsymbol{\alpha}} = \mathbf{X_c}^T\mathbf{x}$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X_c}^T\mathbf{X_c})^{-1}\mathbf{X_c}^T\mathbf{x}$$

This $\hat{\alpha}$ is used to find $\hat{\mathbf{x}}$ given in equation (5.3) as:

$$\hat{\mathbf{x}} = \mathbf{X_c}\hat{\boldsymbol{\alpha}} \tag{5.5}$$

The estimated projection vectors ($\hat{\mathbf{x}}$) are calculated for all classes for the given test image. After that we evaluate the euclidean distance between the test projection vector and estimated projection vectors. The class of the estimated image that gives

minimum distance is said to be the class of the unknown image. The schematic of the LS-PCA technique is shown in Fig. 5.17.

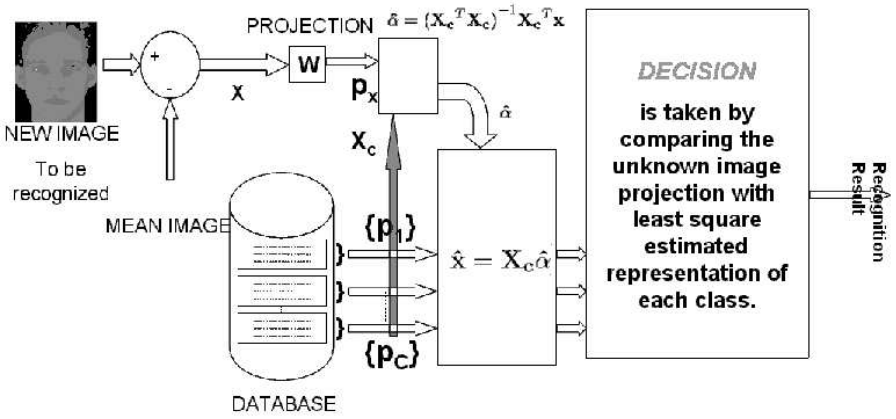


Figure 5.17: Schematic diagram (face recognition using LS-PCA)

5.5.3 Experimental Results

The LS-PCA technique has an advantage over classical PCA technique as it incorporates class information. This results in an increase in the recognition accuracy of the system. As seen earlier, the more the number of eigenvectors used, the better the result. The result is shown in Fig. 5.18. Note that we are using the leave-one-out method for finding the results. The face recognition results for the previous PCA technique are compared with the new LS-PCA results in Fig. 5.19. It can be observed from the two curves (see Fig. 5.19) that simple PCA technique performance is better than the LS-PCA when we use less than 10 eigenvectors. After that the LS-PCA performs better. We investigated and found that this behaviour occurs

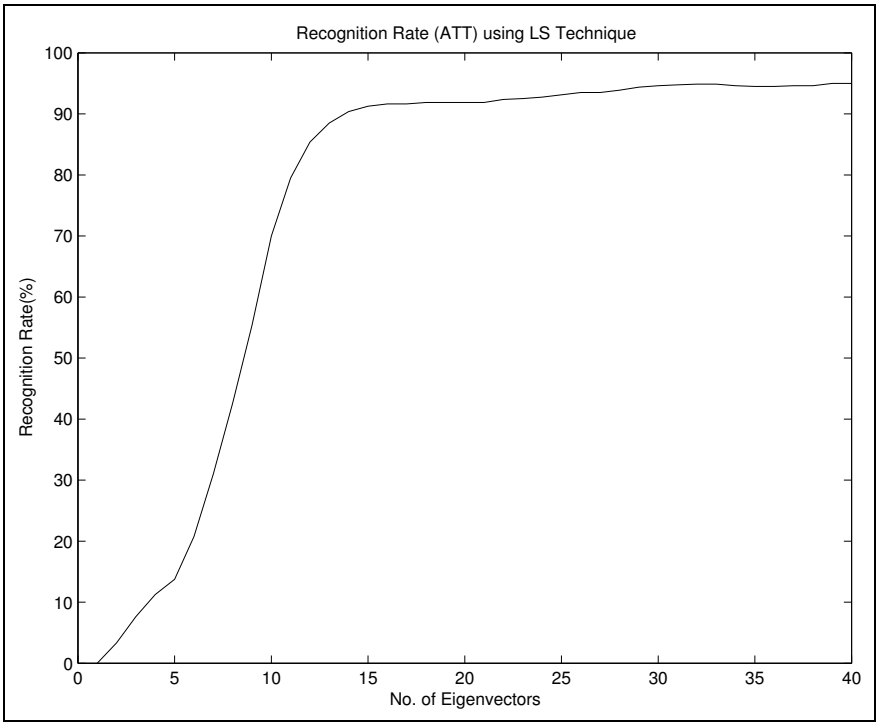


Figure 5.18: Recognition rate, leave-One-Out LSPCA (ATT database)

because of two reasons.

1. Number of images used for comparison

2. Number of Eigenvectors

By number of images used for comparison, we mean that in simple PCA we compare the unknown test image with six images from a class. This results in a higher probability of matching the correct class. While in LS-PCA we are comparing the unknown test image with only one image per class. This deprivation is overcome by using more eigenvectors and thus getting a better estimate of the unknown image.

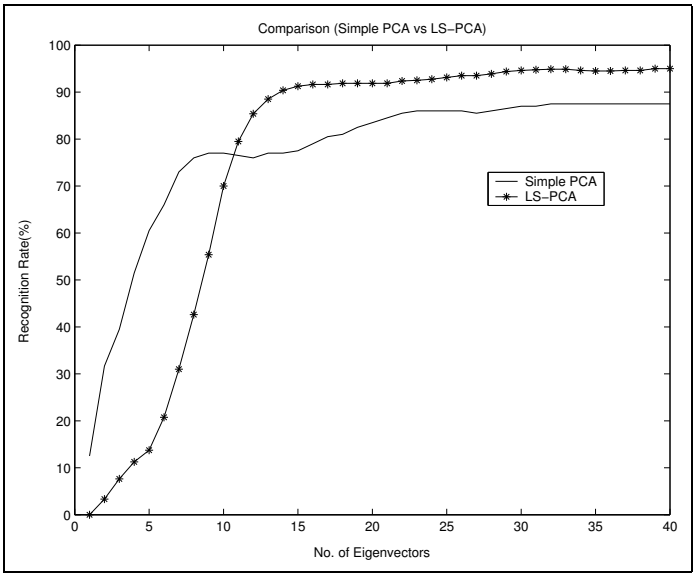


Figure 5.19: Comparison between PCA and LS-PCA results for face images (ATT database)

The LS-PCA when applied over eyes images also give improved performance. The results of simple PCA are compared with the results of LS-PCA in Fig. 5.20 and 5.21

5.6 Summary

In this chapter, we presented two new methods for person identification, one using the face image, and the other using eyes. It was shown that the eyes are good candidates for robust person identification. This followed a discussion of pose variations, including rotation and scale, on face recognition. Some methods were proposed to counter the problems posed by these variations. In the end, the proposed technique of Least Squares PCA was explained. This technique improves the traditional PCA

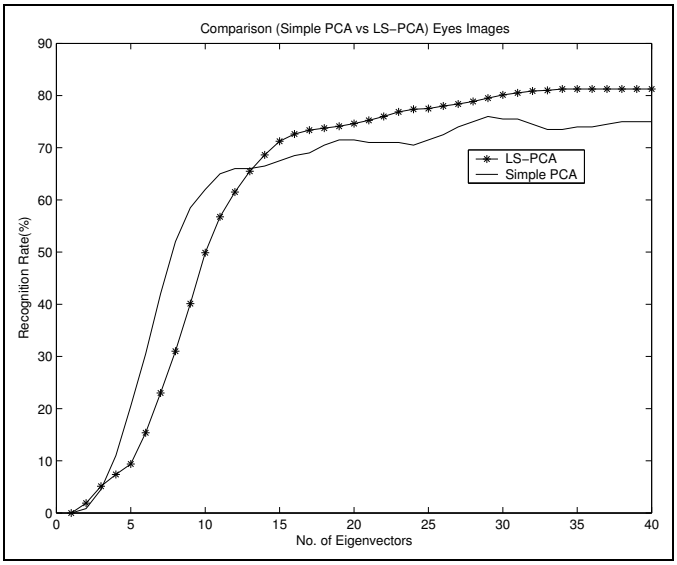


Figure 5.20: Comparison between PCA and LS-PCA techniques when applied for eyes images (ATT database)

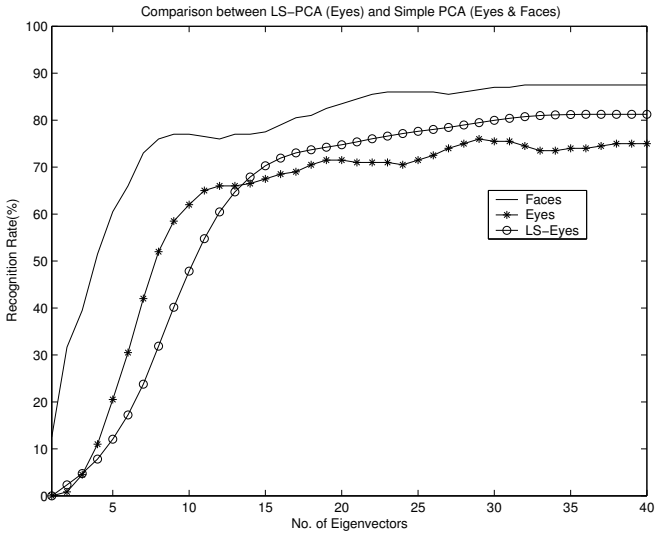


Figure 5.21: LS-PCA eyes results compared to traditional PCA results

by incorporating class information.

Chapter 6

Face Recognition using Multiscale Analysis

6.1 Introduction

The face recognition technique based on PCA (chapter 5) analyzes the input image at only one scale i.e. the finest(original) scale. In multiscale techniques for face recognition, the input images are analyzed at a number of scales. First the input image is decomposed into different subimages using a particular technique. These subimages correspond to different frequency bands hence they are called subband images. These subband images are then used as input to a combined face recognition system. In multiscale face recognition systems, we are able to analyze images at different scales hence improved performance is achieved. We call such approach the

Multiscale PCA (MPCA) approach.

6.2 MPCA System Description

The multiscale decomposition of images can be performed in a number of ways. We are using here the wavelet packet analysis in order to perform such decomposition of input images. The wavelet packet decomposition of images has already been discussed in Sec. 3.2. A simple schematic diagram of the process is shown in Fig. 6.1. In the training stage, each input image is first decomposed into several subband images. Then, from the corresponding subband images of all the training images we form a library. This means if our image decomposition results in N images, we will have N separate libraries. In the testing stage, the input images are first decomposed and each subband image is projected onto the corresponding library that was created during the training stage.

6.3 Experimental Setup

In this section, we will discuss several experiments performed to investigate the multiscale behaviour of images. These experiments were the basis of important findings used later in Sec. 6.5.

The framework for multiscale analysis of images has already been discussed in sections 3.2 and 6.2. Here we use such framework to discuss the experiments that

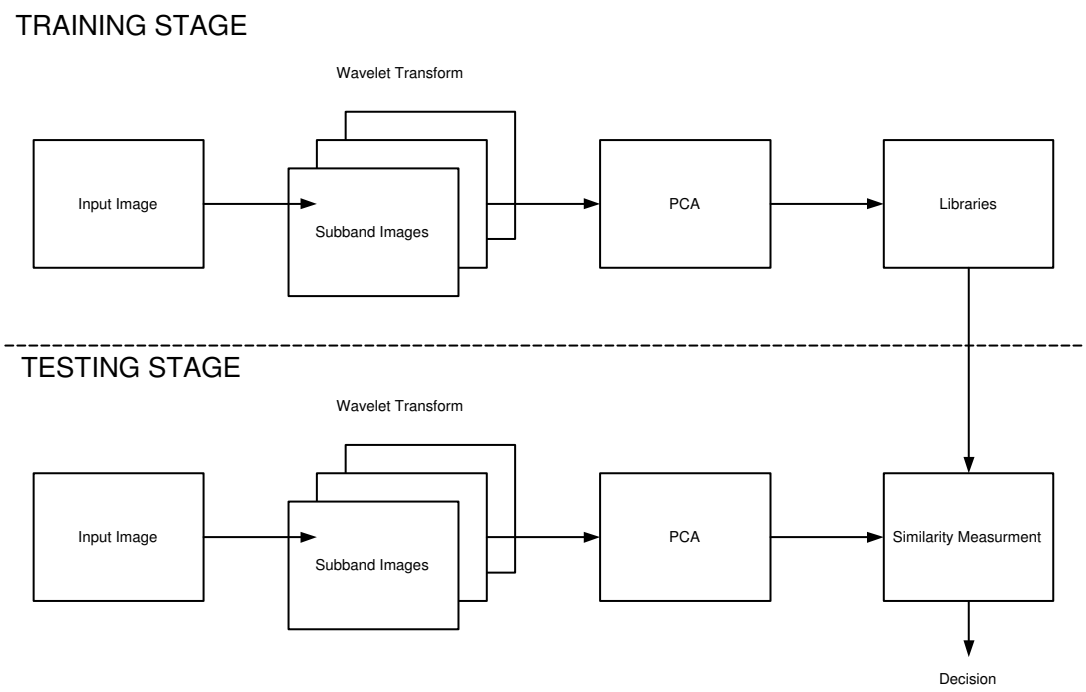


Figure 6.1: MPCA schematic diagram

we performed using subband images. The system developed earlier (chapter 5) is also be used here. The only difference is that instead of supplying simple images as input, we are now supplying the decomposed subband images. There are several parameters involved when we speak about multiscale analysis of an image. First of all, we must select a particular wavelet; in other words, a set of lowpass and highpass filters. Once such filters are selected, we have to determine the level of decomposition and the type of orientation.

Recall that our main objective is to use multiscale information from images to enhance recognition accuracy. This requires a thorough investigation of the multiscale properties of images. To look into the multiscale characteristics of images, we

performed numerous experiments varying the parameters mentioned above. The results were recorded and analyzed to draw a number of conclusions. The experiments were performed to see the effect of following parameters:

- Types of filters (wavelets),
- Orientations,
- Levels, and
- Resolution.

We give below a brief description of the specific experiments we carried.

a. ***Wavelets(filters) used:*** db1, db2, db3, db4, db7, db8, db9, db10, and a pair of filters from the literature. These wavelet filters basically differ in their lengths. The filter length for db1 is two, db2 is four and so on.

b. ***Orientations:*** All possible orientations were considered, i.e. Approximations, Horizontal Details, Vertical Details, and Diagonal Details.

c. ***Levels:*** One to Four levels.

d. ***Databases:*** The ATT, Yale, and AR databases.

e. ***Resolutions:*** Original and their half resolution images of the above mentioned databases were used.

The experiments required that we create new databases with the above specifications. All possible combinations of the above mentioned parameters were considered. For example a database of all 'horizontal details' images at 'level four' was created when 'half resolution ' 'Yale database' was decomposed using 'db4' wavelet. For sake of simplicity we will define a naming convention to call these databases.

<database name><resolution>DB<no. of tr images per class>_<wavelet name>_lvl<level number>_<orientation> where the terms in < > are described in Table 6.1 below. Using the convention our example database may be written as YalehalfDB5db4_lvl4_h.

Term	**Description**	**Possible Values**
<database name>	Name of one of the three main databases	ATT, Yale, and AR
<resolution>	the possible image resolution selected	half, and full
<no. of tr images per class>	The number of images per class used for training the system	4, 5, and 6 (ATT), 5, and 7 (Yale), 4 (AR)
<wavelet name>	Name of the wavelet used for decomposition	db1-db4 (ATT), db1-db4, db7-db9 (Yale), db1 (AR)
<level number>	The decomposition level	1-4
<orientation>	Type of orientation	a (=approximation), h (=horizontal details), v (=vertical details), d (=diagonal details)

Table 6.1: Deatils: Database naming convention

One important issue which should be mentioned here is that both wavelet analysis and wavelet packet analysis were used for the experiments. The wavelet packet

analysis is used to determine the best decomposition tree (Sec. 6.6). For now, when we mean by wavelet decomposition the dyadic wavelet analysis in which only approximation image is decomposed at subsequent levels.

We now discuss our results to understand the effect of changing the different parameters. Some useful information on the databases used is reiterated below in Table 6.2.

Name	**Classes**	**Images per class**	**Training Images**	**Test Images**
ATT	40	10	6	4
Yale	15	11	7/5	4/6
AR	60	10	4	6

Table 6.2: Databases information

6.4 Variations in Recognition Rate: Effects of Filter length, Orientations, Levels, and Image Resolution

6.4.1 Effect of different Filter lengths

For the ATT database, four wavelets (filter sets) were used in the decomposition (db1-db4). The resulting images were supplied to our system described in Sec. 5.2. The results presented in Fig. 6.2 and 6.3 show how the recognition accuracy varies with the change in filter length for the full and half resolution images respectively. In

each graph the recognition rate is plotted against different orientations. Moreover, each graph shows results for all four wavelets for a particular level.

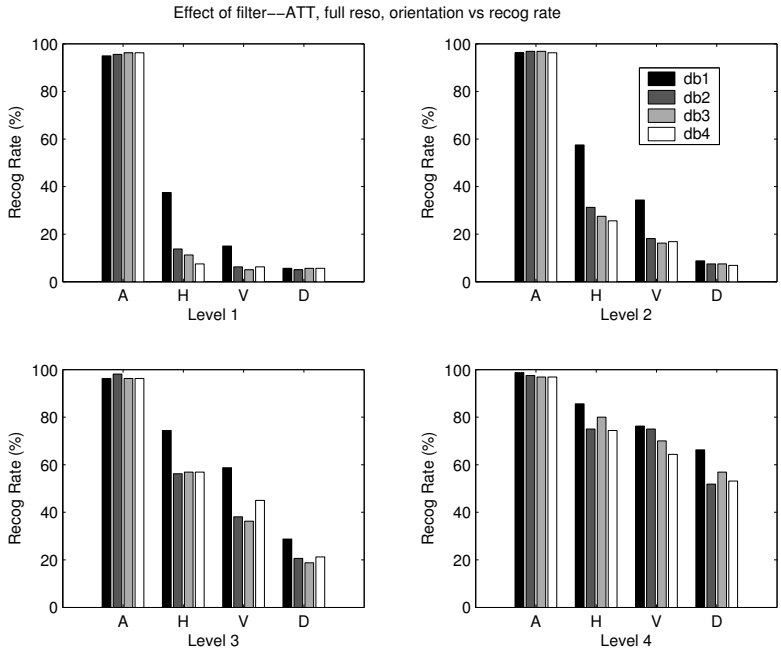


Figure 6.2: Effect of filter lengths on recognition rate (ATT, full resolution)

For levels 1 and 2, the recognition rate drops when the filter length is increased. The best recognition rate is achieved for the level 4 decomposition using the 'db1' wavelet. The results for all the details highlight the importance of 'db1' wavelet which means a filter length of 2. The details images show the edges of an image in different orientations. All edges in a face image span only small number of pixel positions or in other words their thickness is just a few number of pixels. If we use a longer filter to detect edges, the result gives a blurred image with ill-defined edges. While a small filter performs better because it is more likely to detect sharp edges (See Fig. 6.4 for comparison). This explains the reason why we are getting good

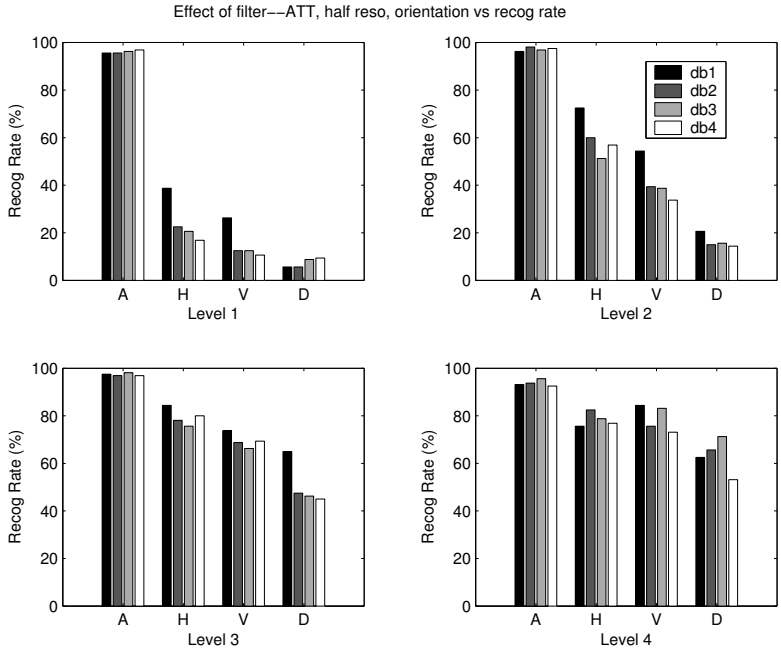


Figure 6.3: Effect of filter lengths on recognition rate (ATT, half resolution)

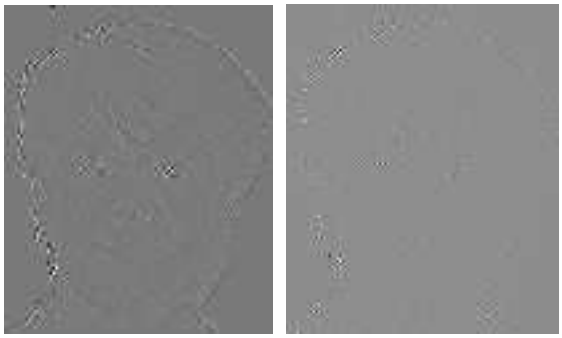
recognition results for subband images generated using small length filters.

Similar results are obtained for the other databases. For example the recognition rates for Yale database when different length filters were used are shown in Fig. 6.5 and 6.6.

In summary, *we observed consistently that the highest recognition rate is obtained for level 4 decomposition with the 'db1' wavelet.*

6.4.2 Effect of Orientations

It can be seen in Figs. 6.7-6.10 that *the best recognition results are obtained for the approximation images.* Though it appears from the results in Figs. 6.7 and



(a) Edges using db1 wavelet

(b) Edges using db4 wavelet

Figure 6.4: Comparison of the edges extracted using different wavelets

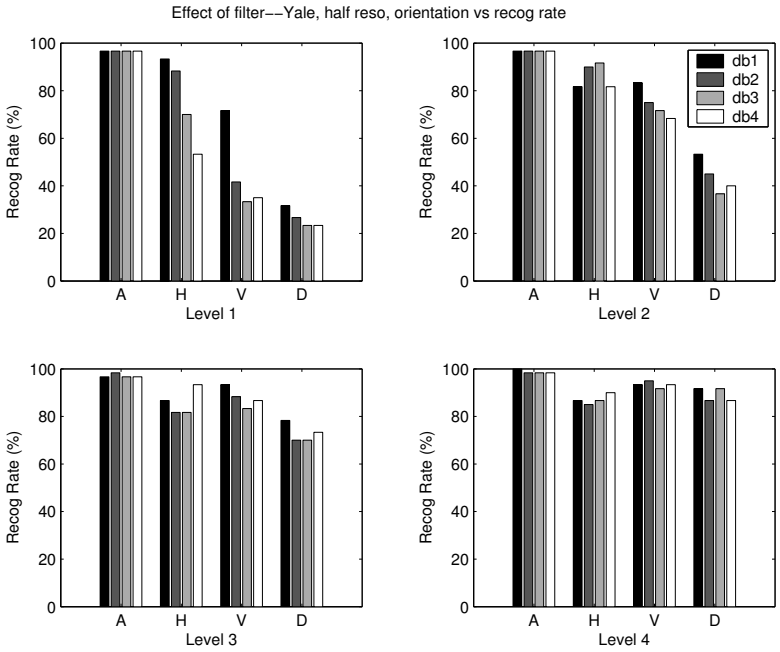


Figure 6.5: Effect of filter lengths on recognition rate (Yale, half resolution)

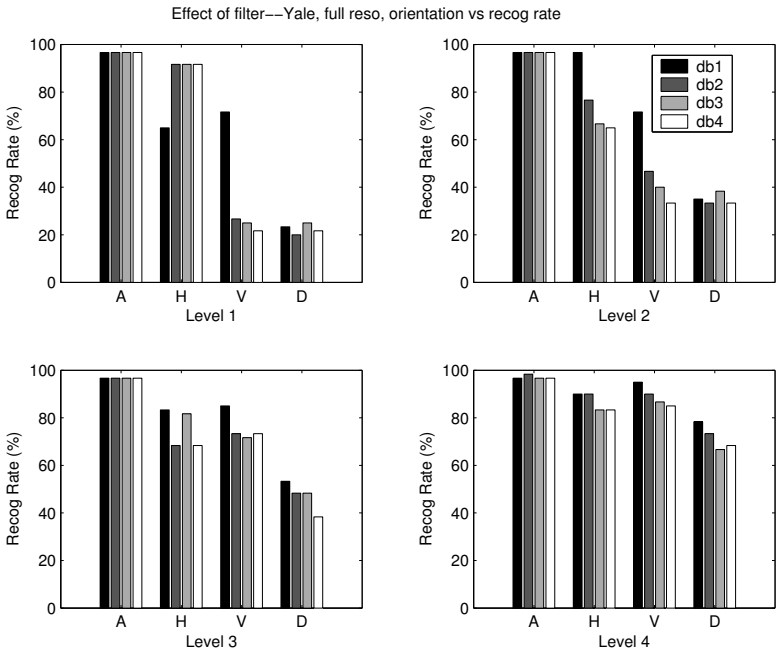


Figure 6.6: Effect of filter lengths on recognition rate (Yale, full resolution)

6.8 that details images also have a significant impact on the recognition accuracy, the effect is much appreciated from the results of Yale database in Figs. 6.9 and 6.10. For the Yale database, we observe that the details play a significant role in recognition with best results for the horizontal details. The horizontal details give better results because of the fact that most human facial features are present in the horizontal direction. For the ATT database, the details also play a significant role but only at higher-level decomposition. As mentioned earlier in Sec. 4.1, the ATT database was scaled to make the dimensions of all face images identical. The process of resizing an image involves lowpass filtering and interpolation. When we try to resize an image to a larger one, we are actually adding some pixels and this requires interpolation to fill in such pixel values. This affects image information,

mostly in the frequency domain. The problem is severe when the new image size is not in integer multiples of the original image size. The edges, which correspond to higher frequencies, in image are also affected badly. On the other hand, if the image is reduced in size we may still need to use interpolation but after lowpass filtering. The lowpass filtering is performed to reduce the effect of aliasing. The frequency content of the original image is again disturbed. Whatever may be the interpolation, nearest neighbour, bilinear, bicubic, the image information is affected. This change in frequency content is the main cause for the results obtained on the ATT database. It can be seen that the recognition rates drop dramatically for the details at lower levels (1 and 2) specially. This behaviour points toward the lack of significant edges in the images.

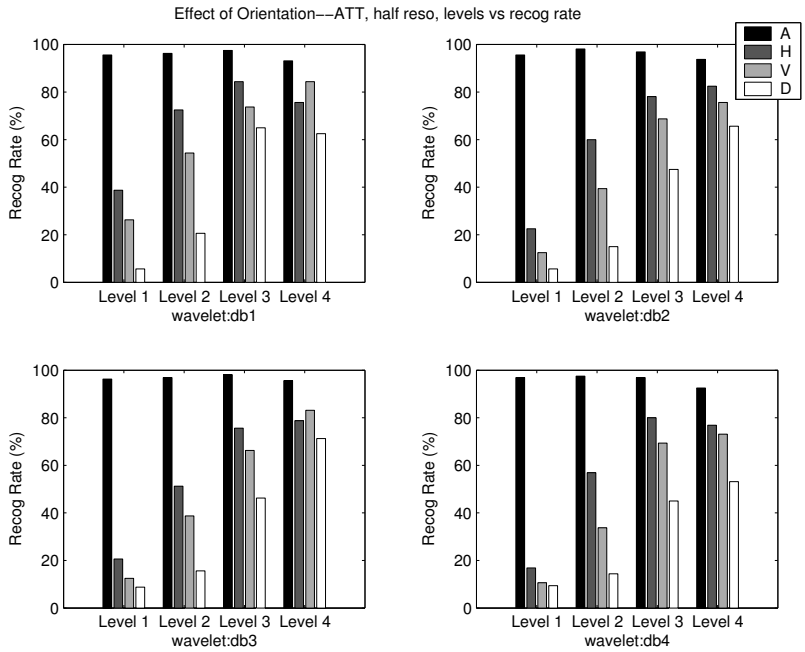


Figure 6.7: Effect of orientation on recognition rate (ATT, half resolution)

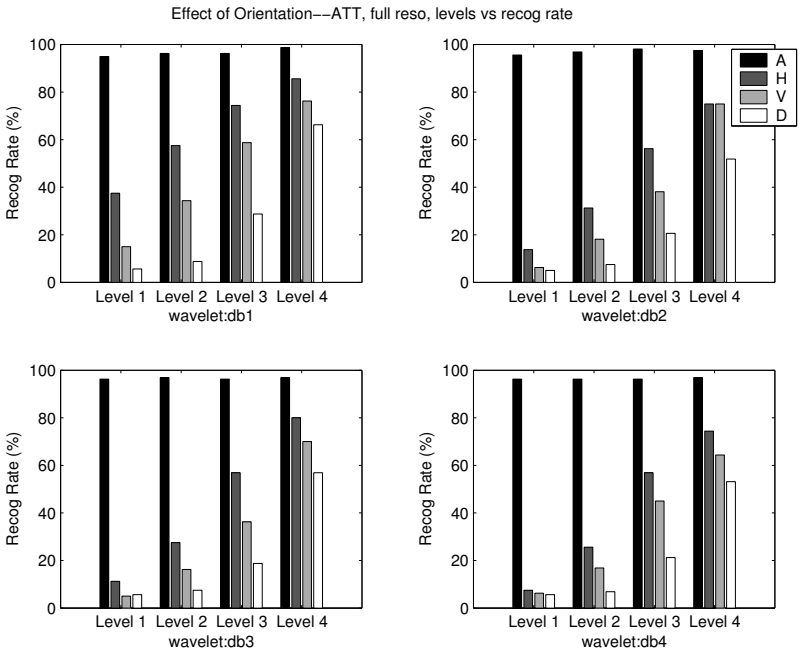


Figure 6.8: Effect of orientation on recognition rate (ATT, full resolution)

6.4.3 Effect of Levels

With increase in decomposition level, the resulting subband image size is reduced. With a decrease in image size the total information content of the image is also reduced. Logically this decrease in information should result in a decrease in recognition accuracy. The results in Figs. 6.11-6.14 show otherwise. *The recognition rate is increasing as we go from level 1 to level 4 (clearer on detail images).* To support our claim and to check this strange behaviour we looked at the entropies of these images. The average entropies of the subband images for different levels are shown in Fig. 6.15. The figure shows that the average entropy increases with increase in

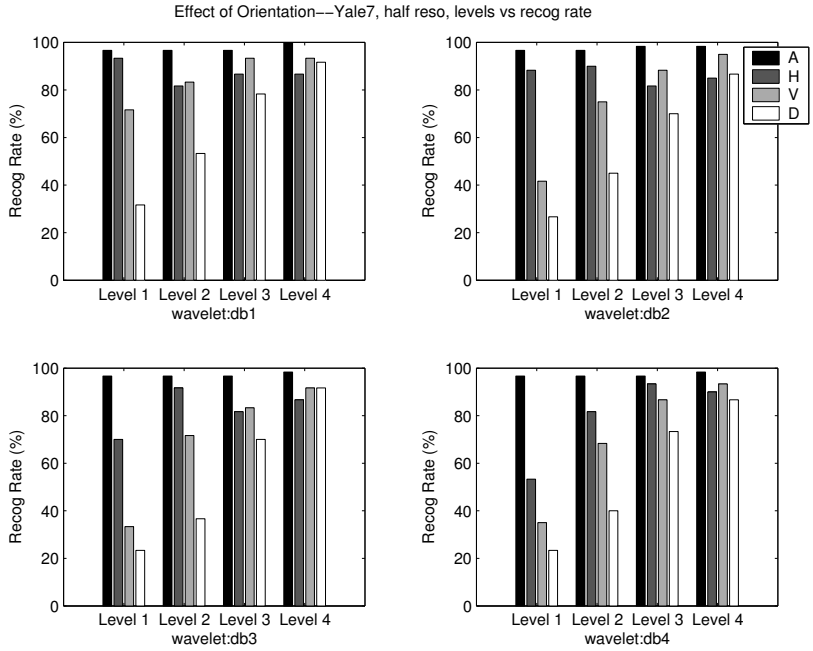


Figure 6.9: Effect of orientation on recognition rate (Yale, half resolution)

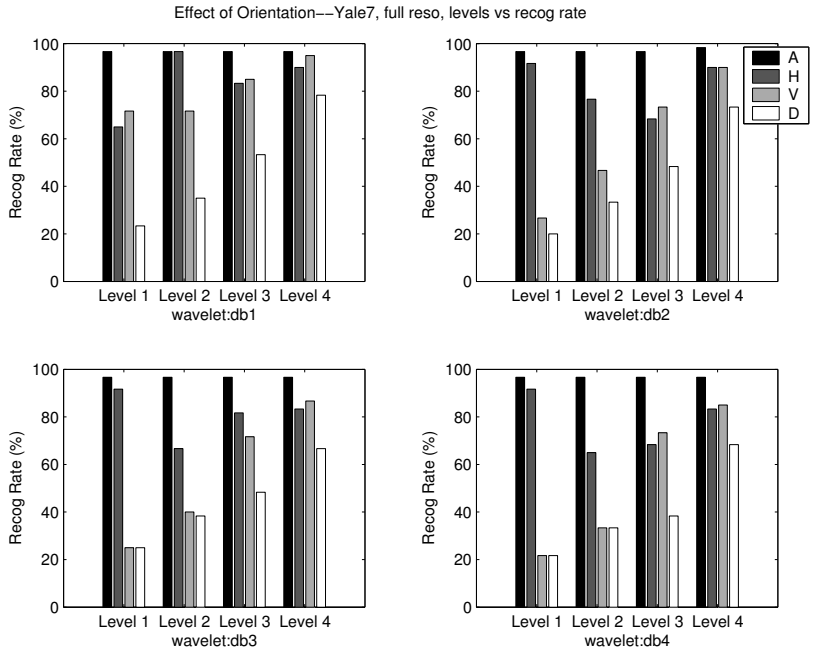


Figure 6.10: Effect of orientation on recognition rate (Yale, full resolution)

decomposition level hence the increase in recognition rate.

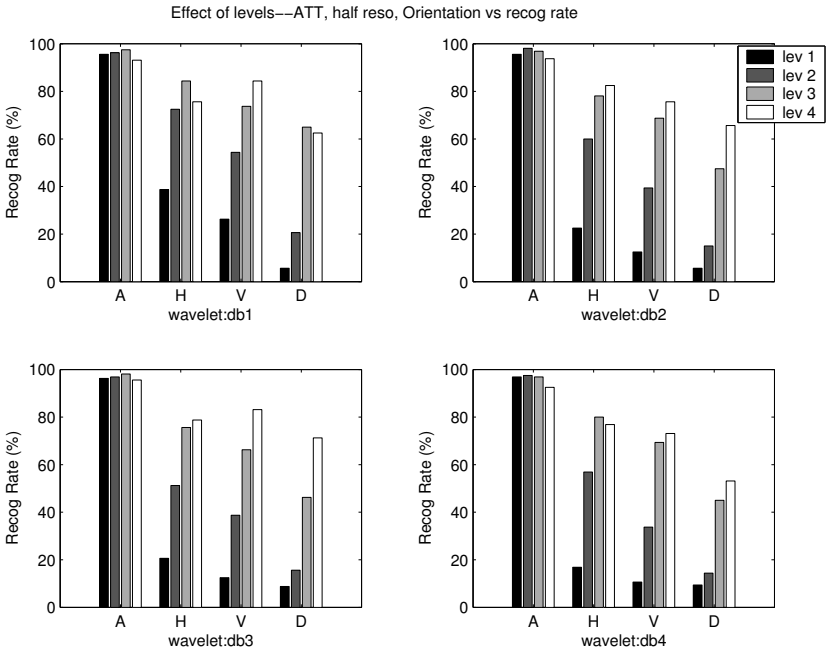


Figure 6.11: Effect of levels on recognition rate (ATT, half resolution)

6.5 Effect of Change in Resolution

The results provided in previous sections are quite self-explanatory and highlight some very important behaviours of images in multiscale environment. They helped us find which of the subbands and at what level is important in the sense of high recognition accuracy.

We found that the results of the same database are different among different resolutions. This meant that the recognition rates are dependent upon the image resolution. We looked into this matter in more detail and found that this behaviour

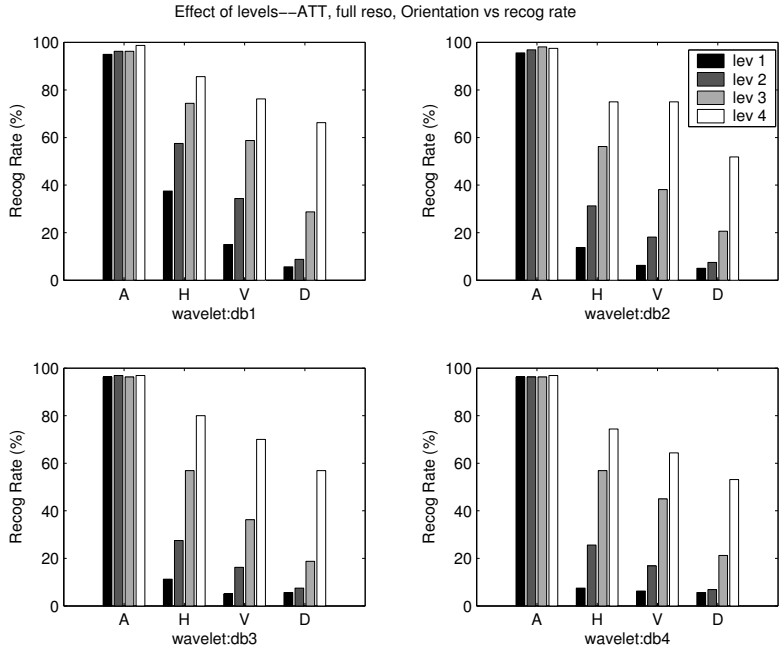


Figure 6.12: Effect of levels on recognition rate (ATT, full resolution)

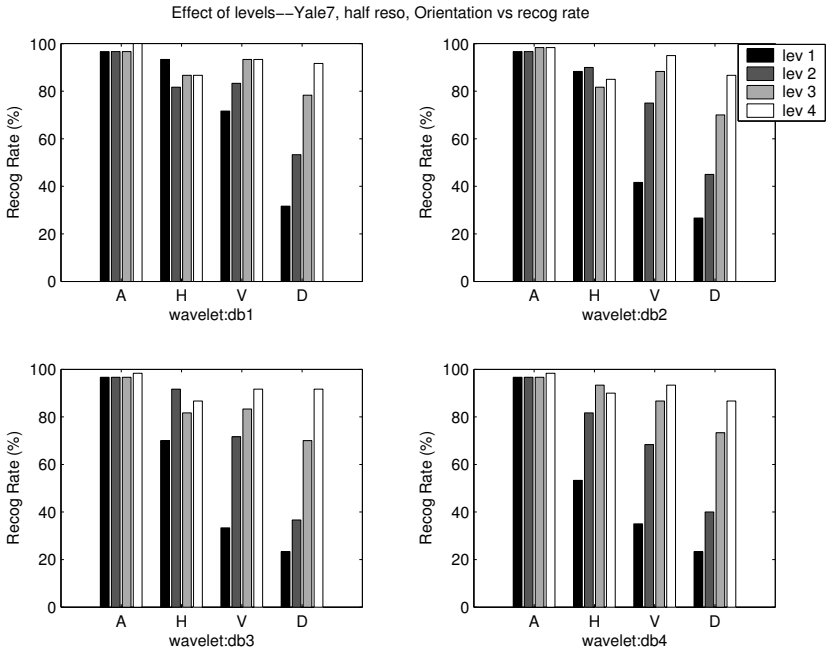


Figure 6.13: Effect of levels on recognition rate (Yale, half resolution)

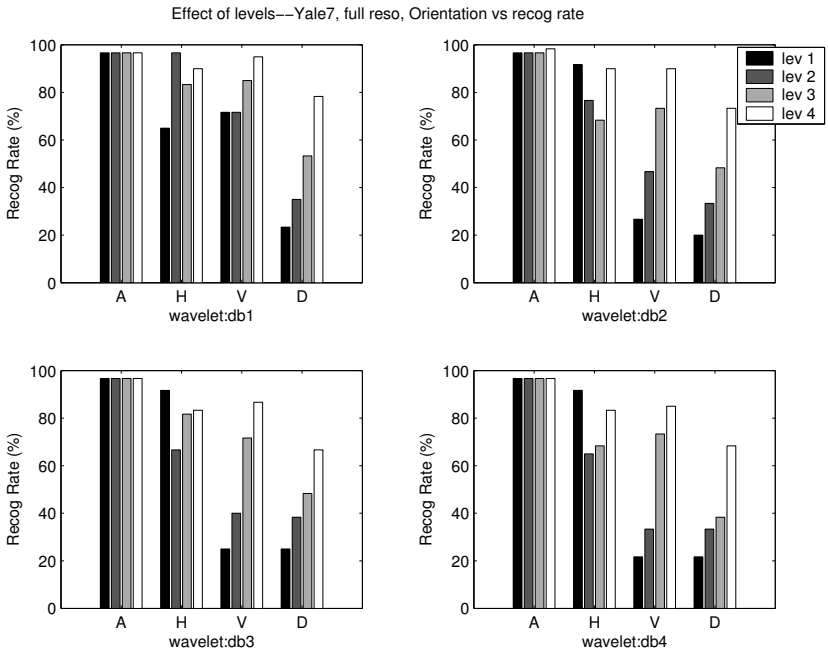


Figure 6.14: Effect of levels on recognition rate (Yale, full resolution)

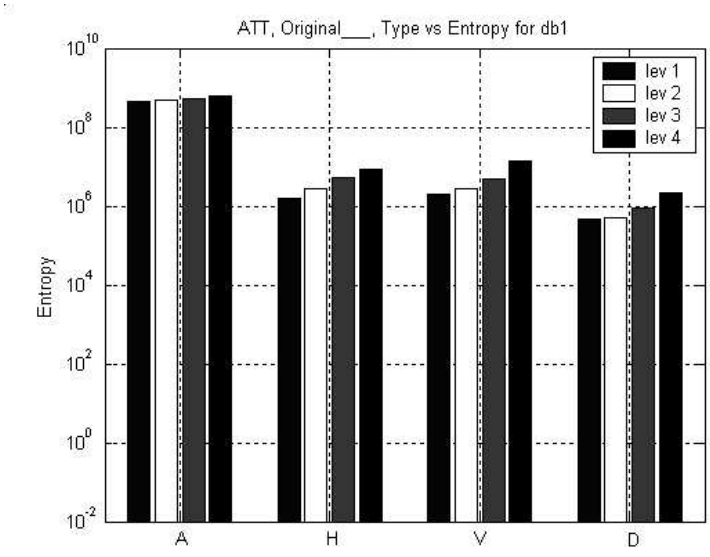


Figure 6.15: Average entropy as a function of subband images at different levels

is actually connected to the decomposition level. If we look carefully, we find that *the recognition rates at level 1 for the half-resolution databases are similar to those at level 2 of full-resolution.* Similarly the results at level 2 and 3 of half-resolution correspond to those at level 3 and 4 respectively of full-resolution. The results do not match exactly but the differences are negligible. For example compare Figs. 6.3 and 6.2. Similarly compare Figs. 6.5 and 6.6 for the Yale database. This behaviour is explained in the following section.

6.5.1 Explanation of the Resolution effect and Nyquist criterion

The form of the results discussed earlier points towards an important aspect; the change in resolution is actually affecting the frequency content of the image. Actually the size of the frequency band is altered. When we decrease the image size we are actually removing some finer details. Hence the frequency band is reduced or in other words some high frequency components are removed.

If we consider an image in its original form with maximum frequency f_{max} then according to the Nyquist sampling theorem, the sampling frequency should be

$$f_s \geq 2f_{max}$$

or in other words if our sampling frequency is f_s, the maximum frequency that

can be present in our image will be equal or less than $f_s/2$ i.e.

$$f_{max} \leq \frac{f_s}{2}$$

This is the upper bound on f_{max}. So in wavelet domain, we get the division of bands as shown in Fig. 6.16. This is because wavelet transform bisects the available bandwidth.

We can easily deduce from here that if the size of the image is changed or in other words if the image resolution is altered, the frequency content is also altered. This helped us explain why the level m half resolution images are behaving as level $m+1$ in full resolution images. When we halved the image size we actually restricted the maximum frequency to $f_{max}/2$. Hence the 'good' subbands that were previously at level m are now at level $m-1$. This can be explained easier using a mathematical formulation.

6.5.2 Mathematical Description

For the purpose of analytical formulation, we will use the wavelet packet decomposition instead of the simple wavelet decomposition.

Lets say that we have an image I of size $N \times N$ with a sampling frequency of f_s and a maximum frequency f_{max}. The wavelet packet decomposition with the different frequency bands is shown in Fig. 6.17. The relation between f_s and f_{max}

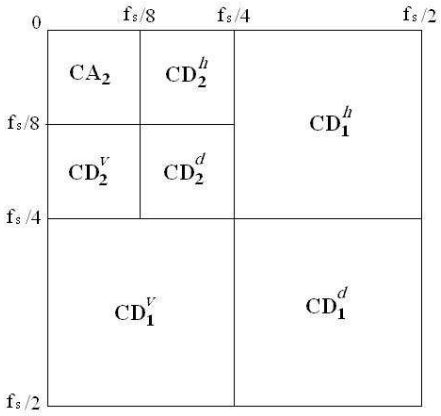


Figure 6.16: Division of bands in wavelet transform

is given below

$$f_{max} \leq \frac{f_s}{2}$$

$$f_s \geq 2f_{max}$$

Lets say that the important subbands or more generally the important frequency band for image I turn out to be from f_s/a to f_s/b with the later having a higher value ($a > b \geq 2$). We now have a new image, resized(downsampled) version of the original I, say I'. The size of I' is $N' \times N'$ and a maximum frequency of

$$f'_{max} = \frac{N'}{N} f_{max}$$

Similarly the new sampling frequency can be deduced to be

$$f'_s \geq 2f'_{max}$$

$$f'_s \geq 2\frac{N'}{N} f_{max}$$

From here we get a relation between the new sampling frequency f_s' and the original image sampling frequency f_s.

$$f_s' = 2\frac{N'}{N} f_{max}$$
$$f_s' = \frac{N'}{N} f_s$$
$$f_s = \frac{N}{N'} f_s'$$

This will help us in determining the "good" frequency band for the new resized image. Obviously the range will remain as from f_s/a to f_s/b. If we replace f_s with the value in terms of f_s' we get

$$\frac{N}{N'}\frac{f_s'}{a} \rightarrow \frac{N}{N'}\frac{f_s'}{b} \tag{6.1}$$

We can now either perform bandpass filtering or find the level number to perform wavelet decomposition.

6.6 The Proposed Model for Subband Image selection

Recall that, given an image, our aim is to find important subbands for face recognition. With all the previous discussion, we showed that there are some subbands that are really important and play an important role in face recognition. But are these subbands same for all images? Are these subbands same for the images from different databases? In fact, we want to know whether the subbands that are important

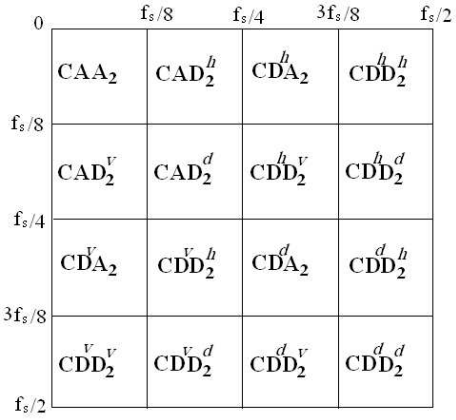


Figure 6.17: Two-level wavelet packet frequency division

for an image from, say, database **A** are also important for database **B**. Obviously, the size will matter but for simplicity lets suppose that the two databases have face images of the same size.

In order to find this we performed experiments on two different databases. These were the Yale and AR databases. Full wavelet packet decomposition of the images from these databases was performed up to level 4. Each image was decomposed into 256 subband images. Then each of the 256 images was used for recognition, resulting in 256 recognition accuracy values. These values are shown graphically in Fig. 6.18 and 6.19 to highlight the importance of some subbands. As can be seen from the colour bar, darker colour stress higher recognition rate. It can be seen that the pattern for the two databases is almost similar and the variations are negligible. Based on this, we can further strengthen our conclusion that *for face images, there are some specific frequency bands that are more useful for recognition. These bands remain the same no matter what the resolution of face images is.* The

only restriction is that no operation is performed which would change the frequency content of the images. Knowledge of these bands and the resolution of image can guide us to determine the wavelet decomposition level that should be used to select the appropriate subband images for recognition.

To progress further and in order to propose a generalized model, we selected the results of the AR database. From the results, we can see that the important subbands lie in region which is highlighted in Fig. 6.20. We can see that for the AR database images the important subbands lie in this region. The range is different for the horizontal and vertical details directions. For the horizontal details it spans the frequency band of $0 - 0.3125 f_s$ while for the vertical the span is larger $0 - 0.3438 f_s$, this is because of the size of the image. The size of the images used for the experiment is 282×244. Note that the *width* of these bands is $0.0625 f_s$. Since we know the resolution (size) of the images, we can now relate these important subbands to any other image. The mathematical description presented in Sec. 6.5.2 can be used to find the important subbands for a new image of size $K \times L$. Using Eq. 6.1 we find the new bands as follows:

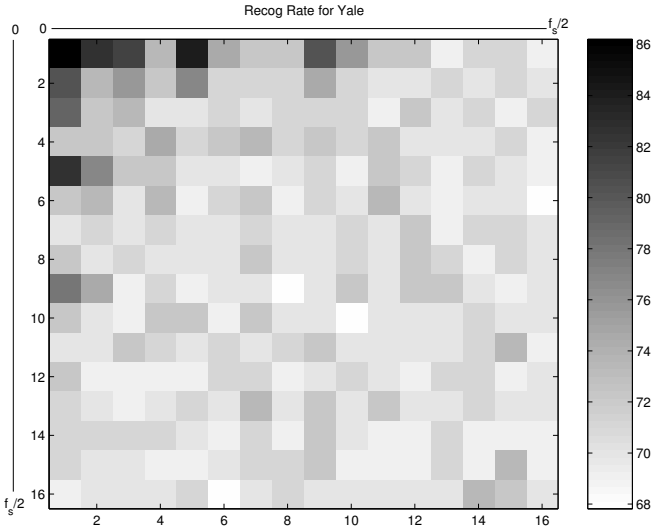


Figure 6.18: Recognition rates at the 4^{th} level wavelet packet decomposition(Yale Database)

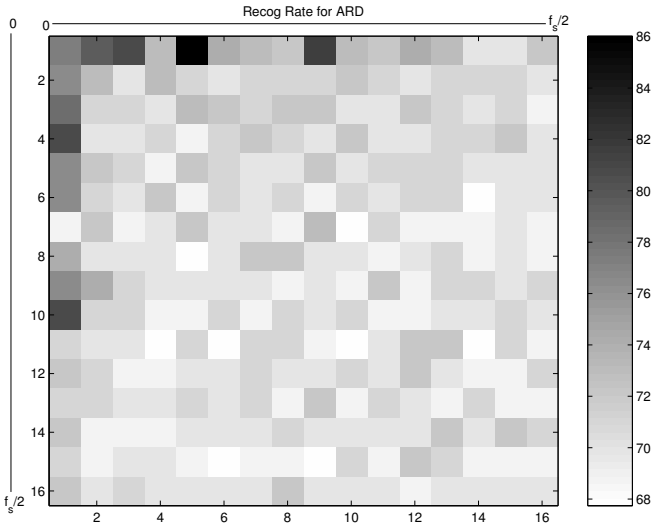


Figure 6.19: Recognition rates at the 4^{th} level wavelet packet decomposition (AR Database)

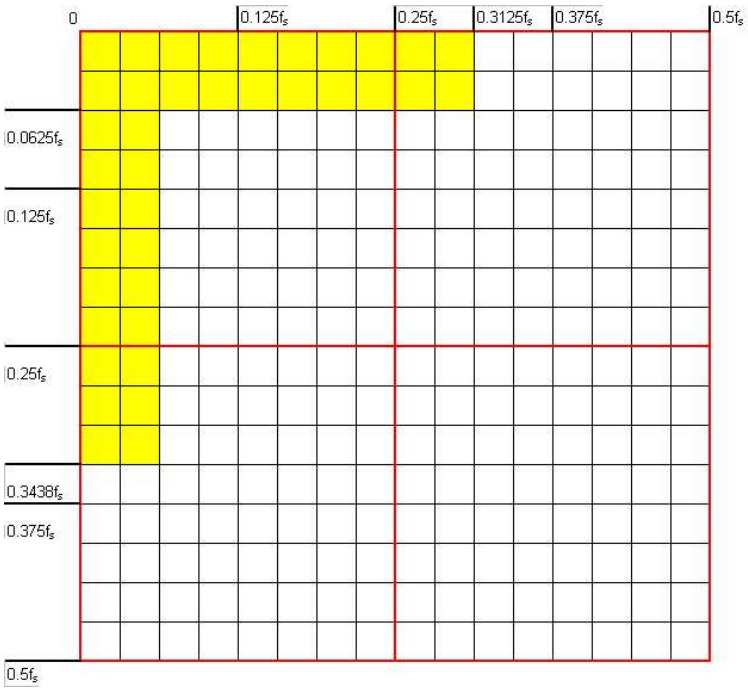


Figure 6.20: Important subbands highlighted, AR

$$\begin{aligned}
\text{Horizontal: } 0 - \frac{282}{K}0.3125f_s' &= 0 - \frac{88.125}{K}f_s' \\
\text{Vertical: } 0 - \frac{244}{L}0.3438f_s' &= 0 - \frac{83.8872}{L}f_s' \\
\text{Width of Horizontal band: } \frac{244}{L}0.0625f_s' &= \frac{15.25}{L}f_s' \\
\text{Width of Vertical band: } \frac{282}{K}0.0625f_s' &= \frac{17.625}{K}f_s'
\end{aligned}$$

So, given any input image, we just put the size of that image in the above equations and find the important subbands for face recognition. For images smaller than 176 in horizontal and 168 in vertical direction, the important band size in the

respective direction will be larger than the total bandwidth of the image. So we use the whole dimension for such cases.

In the next section we propose a new face recognition system using the concept of frequency bands mentioned above.

6.7 The Proposed MPCA Face Recognition System

The fact that we can determine important subbands of an image from the image properties is very helpful in getting a good face recognition system. Instead of using the whole image, we can decompose it and use only the selected subbands. Using this information, we propose a new MPCA-based face recognition system. The flowchart of such face recognition system is shown in Fig. 6.21.

As explained in the figure, we perform PCA on the selected subbands and the individual results for these subbands are combined on the basis of majority voting. Note that the subbands are selected dynamically based on the criterion discussed in previous section. This can be written in the form of an algorithm as given below (Alg. 6.4).

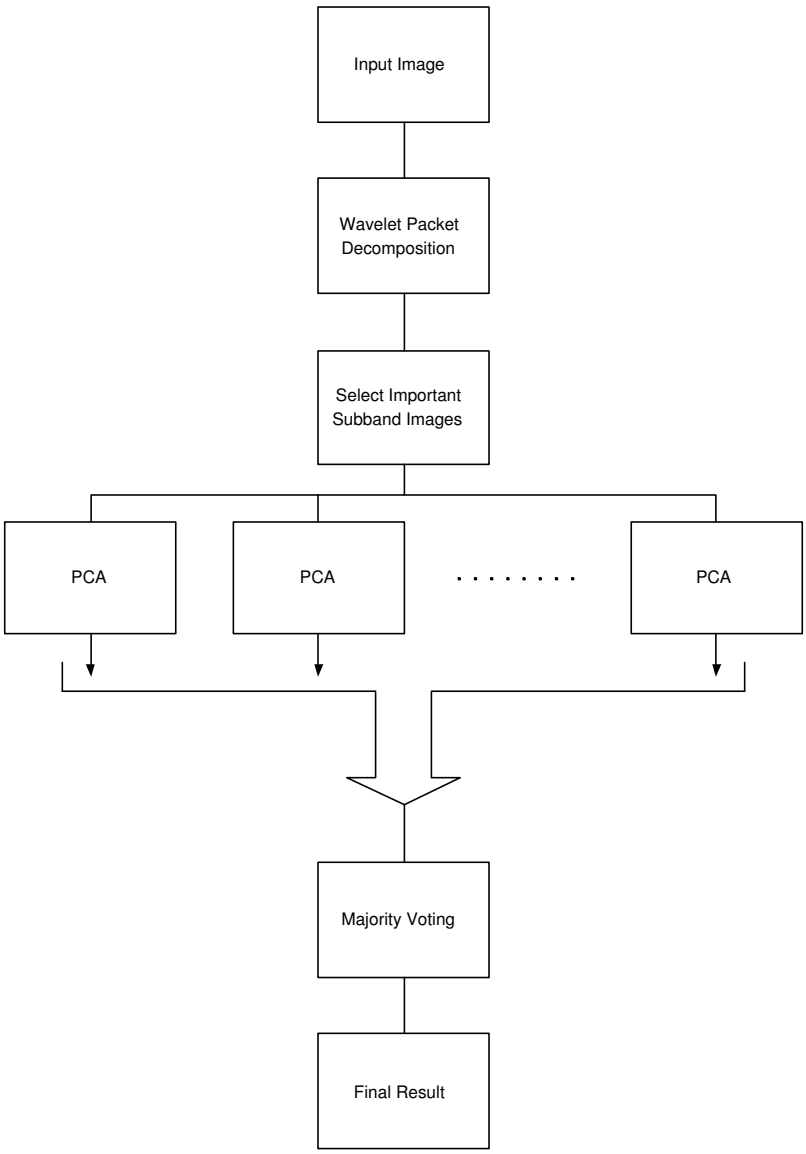


Figure 6.21: Proposed MPCA system schematic diagram

Algorithm 6.4 Proposed MPCA Face Recognition

1: ReadImage I
2: SI = WaveletPacketDec(I)
3: R = FindImageReso(I)
4: IS = CalculateImpSubbands(R)
5: ISI = SelectImpSubbandImages(SI, IS)
6: RR = PCA(ISI)
7: Decision= MajorityVote(RR)

6.7.1 Simulation Results

The algorithm presented above, was implemented and numerous experiments were performed on different databases. The results for the AR database are shown in Figs. 6.22 and 6.23. Fig. 6.22 shows the face recognition results for the simple PCA while those for the proposed MPCA are shown in Fig. 6.23. Both these graphs are superimposed on each other to highlight the superior performance of the proposed MPCA in Fig. 6.24. It is visible that we are able to increase the recognition accuracy by about 5%. It can also be observed that the maximum recognition rate is achieved much quicker for MPCA as compared to the simple PCA technique. Note that for all the graphs shown below we have used three images for training and seven for testing purpose. Note that this technique also has an advantage of a lower computational load. This is because of the fact that we are using a small part of the image instead of using the whole image. For example, in the case of AR database, we are using only 38 subband images as shown in Fig. 6.20. These 38 subband images comprise only 14.84% of the original image and gives very satisfactory results.

6.8 Summary

In this chapter, we presented results of the numerous experiments that were performed to explore the multiscale nature of face images. Important conclusions were drawn from these results which were then used to develop a model for optimal multi-

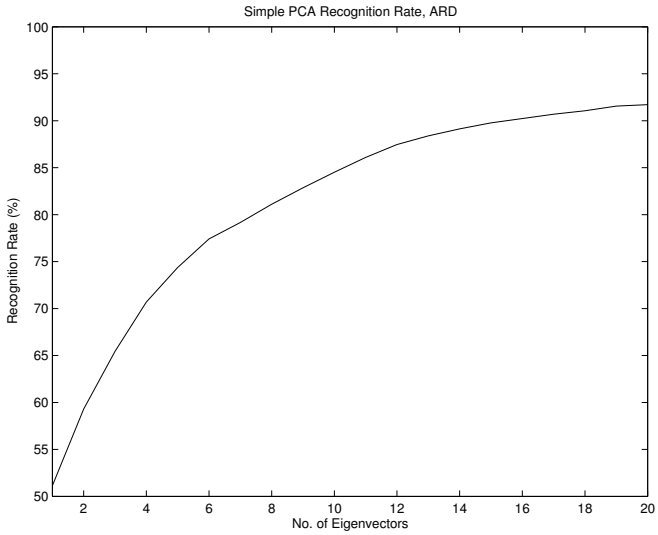


Figure 6.22: Recognition rate, simple PCA (AR database)

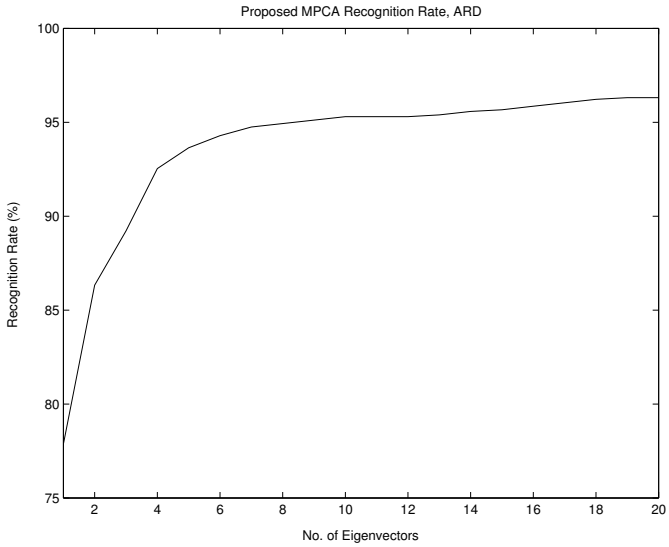


Figure 6.23: Recognition rate, MPCA (AR database)

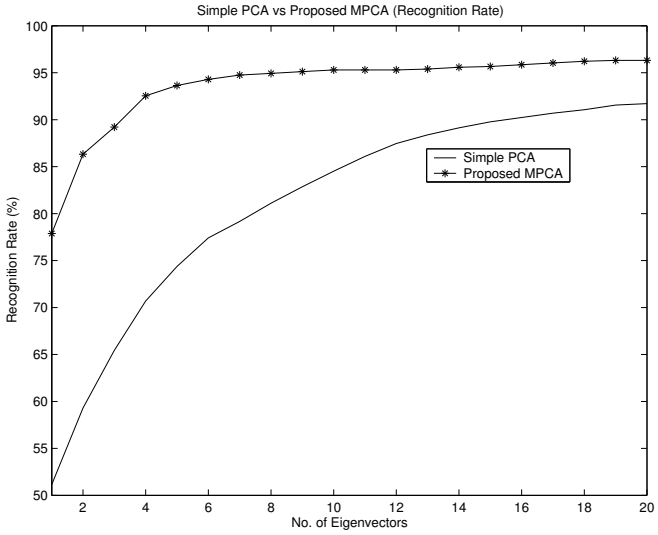


Figure 6.24: Recognition rate, MPCA vs simple PCA (AR database)

scale representation of face images. We then used this model to propose a multiscale face recognition model. The performance of this system was shown to outperform traditional PCA.

Chapter 7

Conclusions and Future Work

7.1 Summary of Results

In this thesis we developed a number of techniques to perform robust face recognition. We developed a framework for face recognition using the multiscale properties of images. The experiments performed to explore the multiscale properties of face images suggest that there exist a particular set of subband images that are useful in face recognition. We state that these subband images span some particular frequency bands which are the same for all face images. We were able to get recognition accuracy much better (5% improvement) than the traditional face recognition system of PCA for the AR database. Since we are using only a fraction of the total image size (14.84% of the original), our technique has an added advantage of low computational complexity.

In addition to the above, which we consider as our major contribution, we also proposed new methods for face/eyes detection, and face recognition. The techniques developed for face and eyes detection can actually be applied to several other problems like face tracking, human-machine interface, emotion identification etc. The Least-Square PCA based face recognition system proposed in this thesis was shown to outperform the classical PCA. This method when used for eyes recognition resulted in improved performance. We were able to achieve an improvement of 8% for face images and 6% for eyes. This method can also work with other facial structures such as nose, ears, and will surely give a better performance.

7.2 Future Work

We developed the model for determining the important subbands for face recognition based on experimental results. We found that the model is a function of input image resolution. A possible extension to the work is to develop a mathematical version of this model. This will require extensive study of the relation between the image resolution and the face recognition technique used.

The thesis presented a translation, and rotation invariant system for face recognition. This was made possible by preprocessing the input images. Such preprocessing stage can be formulated as a mathematical transformation which can be embedded into the proposed rotation and translation invariant system. This will be helpful

when joined with our multiscale face recognition system.

We can further improve the results of multiscale face recognition by performing some kind of classifier combination. Instead of handling subband images separately, and then combining the results, we can combine the information from these subbands to form one common feature vector that represents our input image or use another strategy to combine the recognition results from different subbands, such as the bagging technique.

7.3 Conclusions

Modern face recognition systems are capable of high recognition accuracy. The problem with such systems is that they are vulnerable to pose and scale problems. We proposed some techniques in this thesis to alleviate these problems and make our system rotation and scale invariant. This resulted in significant improvement in performance of face recognition system. Moreover, we showed that by using only a fraction of total image size, and considering the multiscale properties of face images, we were able to develop a robust face recognition system with low computational load.

Bibliography

[1] P.C. Yuen G.C. Feng. Multi-cues eye detection on gray intensity image. *Pattern Recognition*, 34(2001):1033–1046, 2000.

[2] P.C. Yuen G.C. Feng. Variance projection function and its application to eye detection for human face recognition. *Pattern Recognition Lett*, 19:899–906, 1998.

[3] V. Bruce F.F. Soulie H. Wechsler, P.J. Phillips and T.S. Huang. *Face Recognition: From Theory to Applications*, chapter Human Face Perception and Identification, pages 51–72. Eds. Springer-Verlag, Berlin, Germany, 1998.

[4] Burton M. Bruce, V. and N. Dench. What's distinctive about a distinctive face? *Quart. J. Exp. Psych. 47A*, pages 119–141, 1994.

[5] G. Yang and T. S. Huang. Human face detection in complex background. *Pattern Recognition*, 27(1):53–63, 1994.

[6] C. Kotropoulos and I. Pitas. Rule-based face detection in frontal views. *Proc. Intl Conf. Acoustics, Speech and Signal Processing*, 4:2537–2540, 1997.

[7] S.A. Sirohey. Human face segmentation and identification. Technical Report CS-TR-3176, University of Maryland, 1993.

[8] D. Chetverikov and A. Lerch. Multiresolution face detection. *Theoretical Foundations of Computer Vision*, 69:131–140.

[9] E. Petajan H.P. Graf, T. Chen and E. Cosatto. Locating faces and facial parts. *Proc. First Intl Workshop Automatic Face and Gesture Recognition*, pages 41–46, 1995.

[10] M.C. Burl T.K. Leung and P. Perona. Finding faces in cluttered scenes using random labeled graph matching. *Proc. Fifth IEEE Intl Conf. Computer Vision*, pages 637–644, 1995.

[11] K.-C. Yu C.-C. Han, H.-Y.M. Liao and L.-H. Chen. Fast face detection via morphology-based pre-processing. *Proc. Ninth Intl Conf. Image Analysis and Processing*, pages 469–476, 1998.

[12] H. Wechsler B. Takacs. Detection of faces and facial landmarks using iconic filter banks. *Pattern Recognition*, 30(10):1623–1636, 1997.

[13] M.F. Augusteijn and T.L. Skujca. Identification of human faces through texture-based feature recognition and neural network technology. *Proc. IEEE Conf. Neural Networks*, pages 392–398, 1993.

[14] K. Shanmugam R.M. Haralick and I. Dinstein. Texture features for image classification. *IEEE Trans. Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[15] Y. Dai and Y. Nakano. Face-texture model based on sgld and its application in face detection in a color scene. *Pattern Recognition*, 29(6):1007–1017, 1996.

[16] D. Gibbon M. Kocheisen H.P. Graf, E. Cosatto and E. Petajan. Multimodal system for locating heads and faces. *Proc. Second Intl Conf. Automatic Face and Gesture Recognition*, pages 88–93, 1996.

[17] J. Yang and A. Waibel. A real-time face tracker. *Proc. Third Workshop Applications of Computer Vision*, pages 142–147, 1996.

[18] D. Saxe and R. Foulds. Toward robust skin identification in video images. *Proc. Second Intl Conf. Automatic Face and Gesture Recognition*, pages 379–384, 1996.

[19] Y. Raja S. McKenna and S. Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, 17(3/4):223–229, 1998.

[20] K. Sobottka and I. Pitas. Face localization and feature extraction based on shape and color information. *Proc. IEEE Intl Conf. Image Processing,*, pages 483–486, 1996.

[21] A. Goshtasby J. Cai and C. Yu. Detecting human faces in color images. *Proc. 1998 Intl Workshop Multi-Media Database Management Systems*, pages 124–131, 1998.

[22] S.C. Ahn S.-H. Kim, N.-K. Kim and H.-G. Kim. Object oriented face detection using range and color information. *Proc. Third Intl Conf. Automatic Face and Gesture Recognition*, pages 76–81, 1998.

[23] T.S. Jebara and A. Pentland. Parameterized structure from motion for 3d adaptive feedback tracking of faces. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 144–150, 1997.

[24] M.-H. Yang and N. Ahuja. Gaussian mixture model for human skin color and its application in image and video databases. *Proc. SPIE: Storage and Retrieval for Image and Video Databases VII*, 3656:458–466, 1999.

[25] K. Russell T.S. Jebara and A. Pentland. Mixtures of eigenfeatures for real-time structure from texture. *Proc. Sixth IEEE Intl Conf. Computer Vision*, pages 128–135, 1998.

[26] M. Nagao T. Sakai and S. Fujibayashi. Line extraction and pattern detection in a photograph. *Pattern Recognition*, 1:233–248, 1969.

[27] H. Ellis I. Craw and J. Lishman. Automatic extraction of face features. *Pattern Recognition Letters*, 5:183–187, 1987.

[28] K. Wang L. Shen J. Miao, B. Yin and X. Chen. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. *Pattern Recognition*, 32(7):1237–1248, 1999.

[29] P. Hallinan A. Yuille and D. Cohen. Feature extraction from faces using deformable templates. *Intl J. Computer Vision*, 8(2):99–111, 1992.

[30] Y.H. Kwon and N. da Vitoria Lobo. Face detection using templates. *Proc. Intl Conf. Pattern Recognition*, pages 764–767, 1994.

[31] H. Yan K.M. Lam. An analytic-to holistic approach for face recognition based on a single frontal view. *IEEE Trans. Pattern Anal. Mach. Intell*, 20(7):673–686, 1998.

[32] S. Kimura S. Tsuli K. Matsuno, C. Lee. Automatic recognition of human facial expressions. *Proceedings of ICCV*, pages 352–359, 1995.

[33] C.C. Han M.Y. Chern Y.T. Liu S.-H. Jeng, H.Y.M. Liao. Facial feature detection using geometrical face model: an efficient approach. *Pattern Recognition*, 31(3):273–282, 1998.

[34] David J. Kriegman Ming-Hsuan Yang and Narendra Ahuja. Detecting faces in images: A survey. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 24(1):34–58, January 2002.

[35] Erik Hjelmas and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83:236–274, 2001.

[36] H. Yan K.M. Lam. Locating and extracting the eye in human face images. *Pattern Recognition*, 29(5):771–779, 1996.

[37] H. Zhunag X. Xie, R. Sudhakar. On improving eye feature extraction using deformable templates. *Pattern Recognition*, 27:791–799, 1994.

[38] Gao Wen Chen Xu-Lin Shan, Shi-Guang. Facial feature extraction based on facial texture distribution and deformable template. *Journal of Software*, 12(4):570–577, April 2001.

[39] Preteux M. Malicu, F. Tracking facial features in video sequences using a deformable model-based approach. *Proceedings of SPIE*, 2000.

[40] L. Zhang M. Kampmann. Estimation of eye, eyebrow and nose features in videophone sequences. *International Workshop on Very Low Bitrate Video Coding*, pages 101–104, October 1998.

[41] A.M. Tekalp E. Saber. Frontal-view face detection and facial feature extraction using color, shape symmetry based cost functions. *Pattern Recognition Lett*, 19:669–680, 1998.

[42] M. D. Kelly. Visual identification of people by computer. *Stanford AI Project, Stanford, CA.*, (Tech. rep. AI-130), 1970.

[43] T. Kanade. *Computer recognition of human faces.* Birkhauser, Basel, Switzerland, and Stuttgart, Germany, 1973.

[44] L. Sirovich. M. Kirby. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.

[45] M. Turk and Pentland. Eigenfaces for recognition. *J. Cogn. Neurosci*, 3:72–86, 1991.

[46] Baback Moghaddam Alex Pentland and Thad Starner. View-based and modular eigenspaces for face recognition. *Proc. Computer Vision and Pattern Recognition Conf*, pages 1–7, June 1994.

[47] K.-K. Sung and T. Poggio. Example- based learningfor view-based human face detection. *IEEE Transac-tions on Pattern Analysis and Machine Intelligence*, 20:39–51, 1998.

[48] B. G. Haskell A. N. Netravali. Digital pictures. representation and compression, 1988. *New York: Plenum*, 1988.

[49] Y. Kamp H. Bourlard. Auto-association by multilayer perceptrons and singular value decomposition. *Biol. Cybern*, 59:291–294, 1988.

[50] M.A.Kramer. Nonlinear principal component analysis using autoassociativeneural networks. *AIChE*, 37:233–243, 1991.

[51] R. A. Fisher. The statistical utilization of multiple measuremeents. *Ann. Eugen.*, pages 376–386, 1938.

[52] She K. Draper B. A. Beveridge, J. R. and G. H. Givens. A nonparametric statisical comparison of principal component and linear discriminant subspaces for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition.*, 2001.

[53] Avinash C. Kak Aleix M. Martnez. Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, February 2001.

[54] Lades H. M. Bartlett, M. S. and T. Sejnowski. Independent component representation for face recognition. *In Proceedings, SPIE Symposium on Electronic Imaging: Science and Technology.*, pages 528–539, 1998.

[55] E. Hjelmas and J. Wroldsen. Recognizing faces from the eyes only. *Proceedings of the 11th Scandinavian Conference on Image Analysis*, 1999.

[56] Steffans J. Maurer T. Hong H. Elagin E. Neven H. Okada, K. and C. Malsburg. The bochum/usc face recognition system and how it fared in the feret phase iii test. in face recognition: From theory to applications, h. wechsler, p. j. phillips, v. bruce, f. f. soulie, and t. s. huang, eds. springer-verlag, berlin, germany, 186-205.

[57] Chellappa R. Zhao, W. and P. J. Phillips. Subspace linear discriminant analysis for face recognition. *Tech. rep. CAR-TR-914, Center for Automation Research, University of Maryland, College Park, MD.*, 1999.

[58] Kung S. Y. Lin, S. H. and L. J. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Transactions on Neural Networks*, 8:114–132, 1997.

[59] T. Bachmann. Identification of spatially quantized tachistoscopic images of faces: How many pixels does it take to carry identity? *European J. Cog. Psych.*, 3:87–103, 1991.

[60] W. Zhao and R. Chellappa. Illumination-insensitive face recognition using symmetric shape-from-shading. *Conference on Computer Vision and Pattern Recognition.*, pages 286–293, 2000.

[61] Narendra Ahuja Ming-Hsuan Yang, David J. Kriegman. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine In-*

telligence, 24(1):34–58, January 2002.

[62] D. J. Kriegman P. N. Belhumeur. What is the set of images of an object under all possible lighting conditions? *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, page 270, June 18-20 1996.

[63] John Weng Daniel L. Swets. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

[64] A. Krishnaswamy W. Zhao, R. Chellappa. Discriminant analysis of principal components for face recognition. *Proceedings of the 3rd. International Conference on Face and Gesture Recognition*, page 336, April 14-16 1998.

[65] Nastar C. Moghaddam, B. and Pentland. A bayesian similarity measure for direct image matching. *Proceedings, International Conference on Pattern Recognition*, 1996.

[66] I. Craw and P. Cameron. Face recognition by computer. *Proceedings, British Machine Vision Conference*, pages 489–507, 1996.

[67] F. Samaria and S. Young. Hmm based architecture for face identification. *Image Vis. Comput.*, 12:537–583, 1994.

[68] F. Samaria. *Face recognition using hidden markov models.* PhD thesis, University of Cambridge, Cambridge, U.K., 1994.

[69] Norbert Krger Christopher von der Malsburg Laurenz Wiskott, Jean-Marc Fellous. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.

[70] Lades-M. Buhmann, J. and C. Malsburg. Size and distortion invariant object recognition by hierarchical graph matching. *Proceedings, International Joint Conference on Neural Networks.*, pages 411–416, 1990.

[71] J. Buhmann J. Lange C. von der Malsburg R. P. Wurtz W. Konen M. Lades, J. C. Vorbruggen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, March 1993.

[72] B. R. Bakshi. Multiscale pca with application to multivariate statistical process monitoring. *AIChE Journal*, 44(7):1596–1610, 1998.

[73] D.Q. Dai G.C. Feng, P.C. Yuen. Human face recognition using pca on wavelet subband. *Journal of Electronic Imaging*, 9(2):226–233, April 2000.

[74] S. Kawato J. Tang, R. Nakatsu and J. Ohya. A wavelet transform based asker identification system for smart multi-point teleconferences. *Journal of the Visualization Society of Japan*, 20(1):303–306, 2000.

[75] Giorgos Tziritas Christophe Garcia, Giorgos Zikos. A wavelet-based framework for face recognition.

[76] Giorgos Tziritas Christophe Garcia, Giorgos Zikos. Wavelet packet analysis for face recognition. *Image and Vision Computing*, 18:289–297, 2000.

[77] Vytautas Perlibakas. Face recognition using principal component analysis and wavelet packet decomposition. *Informatica 2004*, pages 243–250.

[78] Y. Liu B. Li. When eignfaces are combined with wavelets. *International Journal of Knowledge-based Systems*, 15(5-6):343–347, 2002.

[79] J.T. Chien and C.C. Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1644–1649, 2002.

[80] Moghaddam B. Pentland, A. and T. Starner. View-based and modular eigenspaces for face recognition. *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition.*, 1994.

[81] V. Bruce. Recognizing faces. *Lawrence Erlbaum Associates, London, U.K.*

[82] P. Penev and J. Atick. Local feature analysis: A general statistical theory for objecct representation. *Netw.: Computat. Neural Syst.*, 7:477–500.

[83] D. L. Ruderman. The statistics of natural images. *Netw.: Comput. Neural Syst.*, 5:598–605.

[84] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human face. *J. Opt. Soc. Am.*, pages 519–524, 1987.

[85] Heisele B. Huang, J. and V. Blanz. Component-based face recognition with 3d morphable models. *Proceedings, International Conference on Audio- and Video-Based Person Authentication.*, 2003.

[86] Serre T. Pontil M. Heisele, B. and T. Poggio. Component-based face detection. *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[87] Vetter T. Blanz, V. A morphable model for the synthesis of 3d faces. *Proceedings, SIGGRAPH99.*, pages 187–194.

[88] Vetter T. Blanz, V. Face recognition based on fitting a 3d morphable model. *IEEE Trans. Patt. Anal. Mach. Intell.*, 25:1063–1074, 2003.

[89] R.C.Gonzalez and P.A. Wintz. *Digital Image Processing.* Addison-Wesley, 1987.

[90] B. Noble. *Applied Linear Algebra.* Prentice Hall, 1969.

[91] S.G.Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[92] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transaction on Information Theory*, 5(36):961–1005, 1990.

[93] Ingrid Daubechies. *Ten lectures on Wavelets*. Society for Industrial & Applied Mathematics, May 1992.

[94] G. Strang and T. Nguyen. Wavelets and filter banks. *Wellesley Cambridge Press*, 1996.

[95] J Ivins, J; Porrill. Everything you always wanted to know about snakes (but were afraid to ask). *A I Vision Research Unit, (Department of Psychology), University of Sheffield, England S10 2TP.*, 1993.

[96] A; Terzopoulos D. Kass, M; Witkin. Snakes: Active contour models. *First ICCV*, pages 259–268, 1987.

[97] *AT& T. The Database of Faces.* `http://www.uk.research.att.com/facedatabase.htm`

[98] `http://www.uk.research.att.com/pub/data/att_faces.zip`, *or* `http://www.uk.research.att.com/pub/data/att_faces.tar.Z`.

[99] *The Yale Face Database.* `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`

[100] *The AR Face Database .* `http://rvl1.ecn.purdue.edu/,aleix/aleix_face_DB.html`.

[101] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transaction on Multimedia*, 1(3):264–277, March 1999.

[102] P.J. Phillips W. Zhao, R. Chellapa and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.

Vitae

- Mudassir Masood.
- Born in Karachi, Pakistan on June 17, 1979.
- Received Bachelor of Engineering (B.E) degree in Computer Systems from N.E.D University of Engineering and Technology, Karachi, Pakistan in 2001.
- Joined King Fahd University of Petroleum and Minerals in February 2003.
- Publication: Dr. Mohamed Deriche and Mudassir Masood, 'Eigeneyes Decomposition System for Face Recognition', *(2nd IEEE GCC Conference 2004, Manama, Bahrain).*
- *Publication: Mudassir Masood and Dr. Mohamed Deriche, 'A Least-Squared Eigeneyes Decomposition System for Face Recognition', (8th. Joint Conference on Information Sciences, July 2005, Salt Lake City, USA).*
- *Email: mudassir@kfupm.edu.sa*